

The Observed Squeezed Limit of Cosmological Three-Point Functions

Enrico Pajer,¹ Fabian Schmidt,^{2,3} and Matias Zaldarriaga⁴

¹*Department of Physics, Princeton University, Princeton, NJ 08544, USA*

²*Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*

³*Einstein fellow*

⁴*Institute for Advanced Study, Princeton, NJ 08544, USA*

(Dated: May 7, 2013)

The squeezed limit of the three-point function of cosmological perturbations is a powerful discriminant of different models of the early Universe. We present a conceptually simple and complete framework to relate any primordial bispectrum in this limit to late time observables, such as the CMB temperature bispectrum and the scale-dependent halo bias. We employ a series of convenient coordinate transformations to capture the leading non-linear effects of cosmological perturbation theory on these observables. This makes crucial use of Fermi Normal Coordinates and their conformal generalization, which we introduce here and discuss in detail. As an example, we apply our formalism to standard slow-roll single-field inflation. We show explicitly that Maldacena’s results for the squeezed limits of the scalar bispectrum [proportional to $(n_s - 1)$ in comoving gauge] and the tensor-scalar-scalar bispectrum lead to no deviations from a Gaussian universe, except for projection effects. In particular, the primordial contributions to the squeezed CMB bispectrum and scale dependent halo bias vanish, and there are no “fossil” correlations between long-wavelength tensor perturbations and small-scale perturbations. The contributions to observed correlations are then only due to projection effects such as gravitational lensing and redshift perturbations.

I. INTRODUCTION

In this paper we present a simple and complete framework to derive the late-time physical observables produced by the squeezed limit of primordial three-point functions, in particular those produced during inflation. These observables include the squeezed limit of the CMB bispectrum and the scale-dependent bias for large-scale structure tracers which crucially depends on the bispectrum in the squeezed limit [1–3]. The main obstacle one encounters is the fact that, in order to calculate the bispectrum consistently, one has to work to second order in perturbation theory throughout. We show how this obstacle is overcome when one is interested in the squeezed limit. We take advantage of the fact that different sets of coordinates are convenient in different cosmological epochs and there is a careful choice that makes the computation easy and transparent. In the remainder of the introduction, we present a convenient way to choose coordinates by discussing the different steps that are summarized in Fig. 1.

A first step consists in computing the primordial correlators generated during the phase of *primordial inflation*. For this purpose, coordinates are used that are comoving with respect to some unperturbed universe (see e.g. [4]). A second step consists in describing the local *late-time physical processes*, such as for example the decoupling of photons during recombination or the formation of a dark matter halo, that lead to the generation of some light signal that eventually will reach the Earth. Since these processes take place in regions that are smaller in size than the curvature of the background metric (horizon scale H^{-1}), they are best described in *Fermi Normal Coordinates* (FNC [5]): this is the unique frame (up to three Euler angles) constructed along a timelike geodesic pass-

ing through a point P in which the metric is Minkowski with corrections going as the spatial distance r_F from the central geodesic squared. In an unperturbed FRW Universe, the corrections thus scale as $(Hr_F)^2$, while in the presence of a metric perturbation $h_{\mu\nu}$ the corrections involve second derivatives of $h_{\mu\nu}$ (in particular, a constant or pure gradient metric perturbation is removed on all scales). The last step consists in relating the signals from the point of emission, for example the last scattering surface, or the position of a halo, to the *observables measured on Earth*, where scientists most naturally use FNC centered on the Earth’s world line.

In general, calculating the evolution of the bispectrum of perturbations in any coordinate frame including modes outside the horizon requires a full second-order relativistic calculation in cosmological perturbation theory, including gravity, baryon physics, radiation transfer, etc. This is in general a very complicated problem. However, a drastic simplification takes place when we limit ourselves to a particular configuration of the bispectrum in which two of the momenta, say k_1 and k_2 are much larger than the third, k_3 . This is often referred to as the *squeezed limit* with $k_3 \equiv k_L$ being the *long mode* and $k_1 \sim k_2 \sim k_S$ being the *short modes*.

Then, as described above, we can make use of the fact that in a $k_L/k_S \ll 1$ expansion, the leading and next-to-leading order gravitational effects of the long-wavelength perturbation k_L on the short modes can be removed by transforming to the local FNC frame, as was emphasized in [7–10]. The first non-trivial effect comes from the second spatial and time derivatives of the perturbation, which are suppressed by $\mathcal{O}(k_L^2/k_S^2)$ (see Sec. III for a more precise statement). We will ignore terms of this order in this paper, as they correspond to actual physical interactions, which will need to be treated in detail.

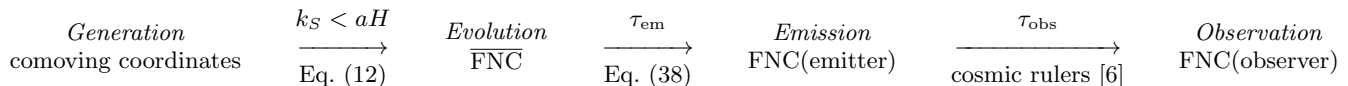


FIG. 1: Sequence of coordinates employed in the computation of observables predicted by a primordial bispectrum in the squeezed limit. The arrows represent the change from one set of coordinates to the next, indicating when this transformation is most conveniently performed and with a reference to the relevant equations. τ_{em} and τ_{obs} denote the conformal times at which the observed photons were emitted and observed, respectively.

However, up to this order, in the local FNC frame we can simply use linear theory. The non-Gaussian correlation will then be captured by averaging over an ensemble of FNC, each constructed to remove a different long mode.

In going from comoving coordinates to FNC, one subtlety arises from the fact that FNC are affected not only by perturbations but also by the expansion of the Universe. Because of this, no matter when the change of coordinates is performed, when using FNC we cannot avoid a period when the evolution at second order in perturbations becomes relevant. To see this, consider an expanding universe with a single long-wavelength perturbation, in addition to short-wavelength modes as in Fig. 2. The metric in standard FNC takes a simple flat space form only on scales smaller than the Hubble scale H^{-1} or the physical wavelength of the long mode a/k_L , whichever is smaller. Therefore for the FNC metric to be valid in a region larger than the short modes, which is what we want to describe, we need $k_S \gg aH$ ($k_S \gg k_L$ being always true by assumption). But as soon as the short modes re-enter the horizon, $k_S \gtrsim aH$, they start evolving, and acquire a non-trivial transfer function. In order to see that the evolution needs to be followed at non-linear order, consider our result for a general squeezed bispectrum Eq. (9). This expression contains terms involving the derivatives with respect to scale and time of the short mode power spectrum, which in general become non-trivial when $k_S \gtrsim aH$ and the short modes start evolving with time. It is clear that these terms cannot be accounted for by just multiplying each of the perturbations with the respective transfer function as *linear* perturbation theory would suggest.

A nice way out of this complication is to use *conformal Fermi Normal Coordinates* (FNC), in which the metric is locally in the FLRW form, i.e. a homogeneous isotropic expanding Universe, rather than in the Minkowski form. These coordinates are valid up to a scale k_L^{-1} , where k_L denotes the wavenumber of the long-wavelength perturbation considered. At late times, the small-wavelength modes are well inside the horizon and have in general been subject to non-linear evolution. Then, transforming from the $\overline{\text{FNC}}$ frame to the usual FNC frame at a given spacetime point simply corresponds to a rescaling of the spatial coordinates.

The final step is to relate quantities defined in the local FNC frame (e.g. temperature of the gas, or halo mass) to the photons actually measured on Earth. This involves the photon propagation (“projection”) effects, such as lensing and redshift-space distortions [6, 11–14]. Here,

we will adopt the “standard ruler” approach [6, 15] since it conveniently encompasses all projection effects for various observables. Specifically, the derivation in Schmidt and Jeong [6] assumes that a ruler, such as a $\xi(r) = \text{const}$ contour, corresponds to a fixed physical spatial scale on a constant-proper-time hypersurface for a comoving observer, equivalent to a fixed spatial scale in FNC at emission.

Finally, at leading order, any *physical* (and necessarily non-gravitational) correlations between long-wavelength modes and small scale modes imprinted at early times and present in $\overline{\text{FNC}}$ then simply add to these projection effects. The framework discussed above, which we summarize in Fig. 1, thus connects the bispectrum calculated in any convenient gauge during inflation with late-time observables such as the CMB bispectrum or the large-scale clustering of tracers. It can be applied to any model, and we will give some examples for single clock inflation and the “fossil” scenarios studied in [16, 17]. We should mention that the importance of FNC in applications to the calculation of inflationary perturbations has been recognized and crucially used previously in the literature, e.g. in deriving various consistency relations [10], in proving that ζ is constant at large distances at all loop orders [9] and in deriving the subleading corrections in k_l/k_s to Maldacena’s consistency condition [8]. The issue of gauge artifacts in the squeezed limit of the bispectrum was also discussed in [18].

The outline of the paper is as follows: in Sec. II we derive some useful formulae to transform correlators from one set of coordinates to another. In Sec. III we introduce $\overline{\text{FNC}}$, and show how the squeezed-limit three-point correlations calculated in the standard way transform to this frame. In Sec. IV we derive how correlations in FNC translate to observables, illustrating that, apart from projection effects such as lensing, the squeezed-limit three-point function in conformal Fermi coordinates is in fact the observed squeezed limit. In other words, if squeezed-limit correlations vanish in $\overline{\text{FNC}}$, then the observed squeezed limit is only due to projection effects. We conclude in Sec. V and leave some technical details to the appendices. App. A and App. B derive the transformation of the two and three point function in the squeezed limit, respectively, as one moves from one set of coordinates to another. In App. C we review the derivation of FNC and discuss their uniqueness in App. D.

II. TRANSFORMATION OF SMALL-SCALE CORRELATIONS AND THE SQUEEZED BISPECTRUM

In this section we present some useful formulae to transform the two-point correlation function, Eq. (4), and three-point function and bispectrum in the squeezed limit, Eq. (9), from any one set of coordinates to another. The details of the derivation are left to appendix App. A and App. B.

Consider a scalar field $\rho(x)$, where $x = (\tau, \mathbf{x})$ denotes the spacetime position. We work in comoving coordinates throughout, so that $x^0 = \tau$ is the conformal time. Under a general coordinate transformation $x \rightarrow x'(x)$, the field transforms as

$$\rho'(x') = \rho(x(x')). \quad (1)$$

Consider a patch on a $\tau' = \text{const}$ surface centered around the position \mathbf{x}'_0 . We would like to derive the correlation $\xi'(\mathbf{r}', \tau')$ within that patch (defined with respect to the mean density $\bar{\rho}'(\tau')$ over the patch) in terms of the two-point correlation function $\xi(\mathbf{r}, \tau)$ of ρ in the unprimed coordinate system. Throughout the paper, we are considering the case where the coordinate transformation varies slowly over the spatial patch, that is where we can expand x in terms of x' as

$$\begin{aligned} x^i(x') &= x_0^i + A^i_j (x'^j - x_0'^j) \\ \tau(x') &= \tau_0 + T_j (x'^j - x_0'^j), \end{aligned} \quad (2)$$

where $x' = (\tau', \mathbf{x}')$, and

$$\begin{aligned} \mathbf{x}_0 &= \mathbf{x}(\mathbf{x}'_0, \tau') \\ \tau_0 &= \tau(\mathbf{x}'_0, \tau') = \tau' + \Delta\tau \\ A^i_j &= \frac{\partial x^i(\mathbf{x}'_0, \tau')}{\partial x'^j} = \delta^i_j - a^i_j. \end{aligned} \quad (3)$$

Note that a^i_j and $\Delta\tau$ are functions of \mathbf{x}'_0 and τ' . Here we have separated A^i_j into a zeroth and first order piece (the sign is chosen so that a^i_j maps x^j to x'^i at linear order). The expansion to linear order in $\mathbf{x}' - \mathbf{x}'_0$ is sufficient, since we will only be interested in Fourier space contributions to the coordinate transformations up to order k_L^2 (see App. A). As shown in App. A, the correlation function $\xi'(\mathbf{r}', \tau')$ of ρ' then becomes

$$\xi'(\mathbf{r}', \tau') = \left[1 - a_{ij} r'^i \partial_{\tau'}^j + \Delta\tau \partial_\tau \right] \xi(\mathbf{r}', \tau'). \quad (4)$$

The dependence on x'^j of the time shift Eq. (2) does not contribute at leading order.

We now turn to the squeezed limit of the three-point function $\langle X'(\mathbf{x}'_3) \delta'(\mathbf{x}'_1) \delta'(\mathbf{x}'_2) \rangle$, where $\delta = \delta\rho/\bar{\rho}$ denotes the fractional perturbation to ρ (the results are identical when considering $\delta\rho$ instead of δ). This limit corresponds to the case where $|\mathbf{x}_3 - \mathbf{x}_1| \gg |\mathbf{x}_1 - \mathbf{x}_2|$. Here, X stands for any other field (X does not have to be a

scalar, although we suppress any tensor indices). Strictly speaking, X is to be understood as coarse-grained on a scale $R > |\mathbf{x}_2 - \mathbf{x}_1|$. In the following, we will drop the primes on coordinates for clarity, since we will not employ the unprimed coordinates anymore. Further, we adopt the notation $X'(\mathbf{x}) = X(\mathbf{x})$, that is, the long-wavelength (coarse-grained) field X is not modified under the small-scale coordinate transformation within the patch. In the squeezed limit, the three-point function describes the modulation of the local two-point function $\xi'(|\mathbf{x}_1 - \mathbf{x}_2|; \mathbf{x}_0)$ at the location $\mathbf{x}_0 = (\mathbf{x}_1 + \mathbf{x}_2)/2$ by the long-wavelength field X evaluated at the distant point \mathbf{x}_3 :

$$\begin{aligned} \langle X'(\mathbf{x}_3) \delta'(\mathbf{x}_1) \delta'(\mathbf{x}_2) \rangle &\stackrel{\text{squeezed}}{=} \\ \langle X'(\mathbf{x}_3) \xi'(\mathbf{x}_1 - \mathbf{x}_2; \tau) |_{X'(\mathbf{x}_0)} \rangle. \end{aligned} \quad (5)$$

The precise location of the point \mathbf{x}_0 in fact does not matter in the squeezed limit (as we prove in App. B), but we have chosen the midpoint as the most natural choice. We can now express $\langle X' \delta' \delta' \rangle$ in terms of the three-point function in unprimed coordinates $\langle X \delta \delta \rangle$ and contributions from the transformation of small-scale correlations [Eq. (4)]. This yields

$$\begin{aligned} \langle X'(\mathbf{x}_3) \delta'(\mathbf{x}_1) \delta'(\mathbf{x}_2) \rangle &\stackrel{\text{squeezed}}{=} \left[1 - \xi_{X a_{ij}}(|\mathbf{x}_3 - \mathbf{x}_0|) r^i \partial_\tau^j \right. \\ &\quad \left. + \xi_{X \Delta\tau}(|\mathbf{x}_3 - \mathbf{x}_0|) \partial_\tau \right] \xi(\mathbf{x}_1 - \mathbf{x}_2; \tau) \\ &\quad + \langle X(\mathbf{x}_3) \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \rangle, \end{aligned} \quad (6)$$

where $\xi_{XY}(r)$ denotes the cross-correlation between X and $Y = a_{ij}, \Delta\tau$. This expression gives the squeezed-limit three-point function of X' and δ' (in the *primed* coordinates) in terms of derivatives of $\xi(\mathbf{r}, \tau)$ and the three-point function of X and δ in the *unprimed* coordinate frame. As shown in App. B, we can derive an analogous expression in Fourier space, corresponding to the squeezed limit of the bispectrum:

$$\begin{aligned} B_{X' \delta' \delta'}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) &\stackrel{\text{squeezed}}{=} \\ &\left[P_{X a_{ij}}(k_L) \left(\delta_{ij} + \frac{k_S^i k_S^j}{k_S^2} \frac{d}{d \ln k_S} \right) P(k_S, \tau) \right. \\ &\quad \left. + P_{X \Delta\tau}(k_L) \frac{\partial}{\partial \tau} P(k_S, \tau) \right]_{\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_2/2} \\ &\quad + B_{X \delta \delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \end{aligned} \quad (7)$$

As before, $P_{XY}(k)$ denote cross-power spectra between X and Y , while $P(k, \tau)$ denotes the power spectrum of ρ . We can now further decompose a_{ij} as

$$a_{ij} = \frac{1}{n_D} a \delta_{ij} + a_{ij}^T, \quad (8)$$

where n_D is the dimensionality of the space in which we define the correlations, and a_{ij}^T is traceless. Allowing

$n_D \neq 3$, in particular $n_D = 2$ will become useful when dealing with projected observables such as the CMB. We then obtain

$$B_{X'\delta'\delta'}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) \stackrel{\text{squeezed}}{=} \left[P_{Xa}(k_L) P(k_S, \tau) \frac{d \ln(k_S^{n_D} P(k_S, \tau))}{d \ln k_S} + P_{Xa_{ij}^T}(k_L) \frac{k_S^i k_S^j}{k_S^2} \frac{dP(k_S, \tau)}{d \ln k_S} + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P(k_S, \tau) \right]_{\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2} + B_{X\delta\delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \quad (9)$$

Note that the contribution of the trace of a_{ij} scales as the logarithmic derivative of $k^{n_D} P(k)$, while the trace-free component a_{ij}^T couples to the logarithmic derivative of $P(k)$ itself. Instead of writing the bispectrum in terms of k_S , one can also expand in k_L/k_1 , yielding zeroth order terms (with $k_S \rightarrow k_1$) and first-order terms $\propto k_L/k_1$. The corresponding expression is given in App. B.

The results of this section are useful when, say, $B_{X\delta\delta}$ is easy to calculate, while $B_{X'\delta'\delta'}$ is more readily related to observations. This will be the case we encounter below.

III. CONFORMAL FERMI NORMAL COORDINATES

Our goal is to derive what the observable consequences are (as opposed to coordinate artifacts) of a given primordial bispectrum. For this purpose we follow the steps in Fig. 1, starting with conformal Fermi Normal Coordinates (FNC). As we discussed in the introduction, if we want to avoid studying second order cosmological perturbation theory, it is not possible to directly connect the statistics of perturbations in the standard FNC frame of galaxies or the CMB (at the time when the photons we observe were emitted) with the statistics calculated when all perturbations are superhorizon. The reason is that the perturbative corrections to the Fermi frame metric are of order $(Hx_F)^2$, where x_F are the physical FNC coordinates, so that the metric in FNC is not approximately flat on the scales of superhorizon perturbations. On the other hand, if we were to move from the global coordinates to FNC at a later time when the short wavelength modes have long re-entered the horizon, we would have to take into account their evolution after horizon entry at second order in the global coordinates.

To overcome this obstacle, in this section we introduce a new coordinate frame which is valid throughout inflation and the hot big bang era (assuming nothing dramatic happens to superhorizon perturbations during this time), and further connects easily with the FNC frame at late times. This provides a well-defined framework for studying observables related to primordial non-Gaussianities in the squeezed limit. To visualize the advantage of FNC,

consider Fig. 2. Each one of the four panels shows the Hubble scale (solid circle) and the region of validity of the FNC (dashed line) at a different moment in time. The long (red line) and short (blue line) modes are also shown for comparison. One can immediately appreciate the fact that from very early time when the long mode is still inside (or about to leave) the horizon all the way until late time when the long mode has re-entered the horizon, the region of validity of $\overline{\text{FNC}}$ amply encompasses the short modes. This is in contrast with standard FNC for which the presence of the Hubble scale makes this impossible. The primordial correlators, typically computed in comoving coordinates, take a simple form once both the short and long modes have left the horizon and cease to evolve. Any time after that moment and before the small-scale modes reenter the horizon constitutes a good time to transform from global conformal coordinates to $\overline{\text{FNC}}$.

Let us assume we are given the primordial correlations during inflation in some gauge, where the metric is written as

$$ds^2 = a^2(\tau) [\eta_{\mu\nu} + h_{\mu\nu}] dx^\mu dx^\nu = a^2(\tau) \bar{g}_{\mu\nu} dx^\mu dx^\nu. \quad (10)$$

Here we have denoted the conformal metric with a bar. In the absence of perturbations, $h_{\mu\nu} = 0$ and $\bar{g}_{\mu\nu} = \eta_{\mu\nu}$. We will work up to linear order in metric perturbations.

As in the usual Fermi normal coordinate construction, we consider a central timelike geodesic of a comoving observer. However, instead of constructing the FNC with respect to $g_{\mu\nu}$, we construct *conformal Fermi normal coordinates* ($\overline{\text{FNC}}$) with respect to $\bar{g}_{\mu\nu}$. That is, within a region around the central geodesic, the metric in these coordinates is approximately $\bar{g}_{\mu\nu}^F = a^2(\tau_F) \eta_{\mu\nu}$. The corrections to this metric, which determine the size of this region which we will call “ $\overline{\text{FNC}}$ patch”, are given by second derivatives of the metric perturbations. Correspondingly, the size of the $\overline{\text{FNC}}$ patch is linked to the wavelength of the metric perturbations considered. In the following, we will be interested in (long-wavelength) perturbations with wavenumber k_L , so that the size of the $\overline{\text{FNC}}$ patch is of order k_L^{-1} .

The transformation from coordinates x^μ to $\overline{\text{FNC}}$, for a patch centered at the spatial origin at time $x^0 = \tau_F$ is given by (see App. C2)

$$\begin{aligned} \bar{x}_F^0(x^\alpha) &= x^0 - \frac{1}{2} \int_0^{x^0} h_{00}(\tau) d\tau - (v_i + h_{0i}) x^i \\ &\quad - \frac{1}{4} [h_{0i,j} + h_{0j,i} + h'_{ij}] x^i x^j + \mathcal{O}[(x^i)^3] \quad (11) \\ \bar{x}_F^k(x^\alpha) &= x^k - v^k (x^0 - \tau_F) + \frac{1}{2} h^k_{i} x^i \\ &\quad + \frac{1}{4} [h^k_{i,j} + h^k_{j,i} - h^k_{ij}] x^i x^j + \mathcal{O}[(x^i)^3]. \quad (12) \end{aligned}$$

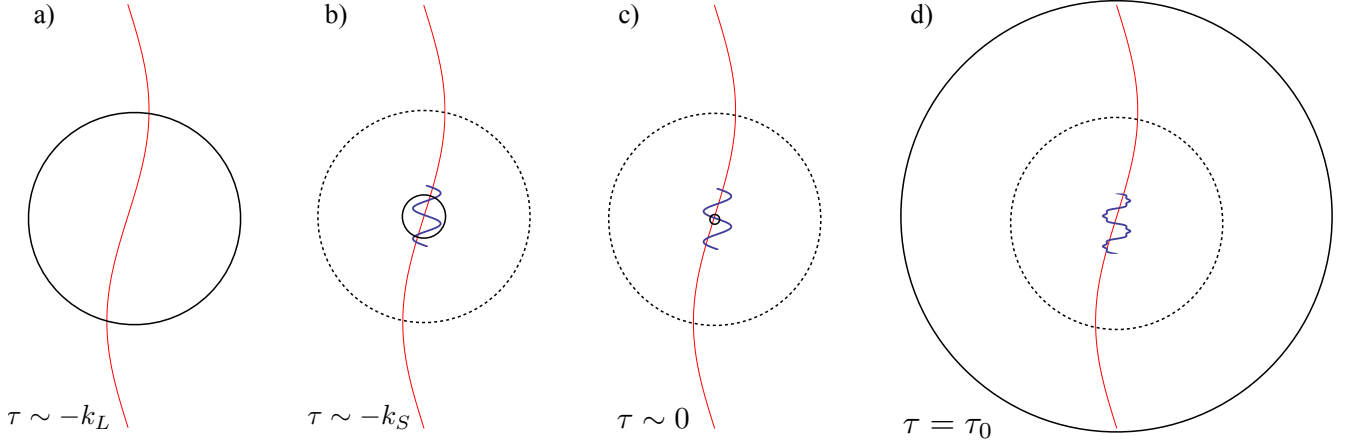


FIG. 2: Illustration of the $\overline{\text{FNC}}$ patch throughout cosmic history, in comoving units. The wavy lines indicate perturbations. The solid circles denote the comoving horizon $1/aH$, which coincides with the size of the usual Fermi coordinate patch during inflation. The dashed circles denote the size of the $\overline{\text{FNC}}$ patch, i.e. the region within which the metric is of the form $\bar{g}_{\mu\nu}^F = a^2 \eta_{\mu\nu}$ with small corrections. a) During inflation, when the long wavelength perturbation $h_{ij}(\mathbf{k}_L)$ is generated (i.e. leaves the horizon), thin red line. b) Later on during inflation; $h_{ij}(\mathbf{k}_L)$ is far outside the horizon when the short wavelength modes $\vartheta(\mathbf{k})$ are generated (thick blue line). The perturbations are generated well within the FNC patch corresponding to the long-wavelength mode. c) Near the end of inflation. All perturbations are far outside the horizon. Nevertheless, the small-scale modes are still well within the $\overline{\text{FNC}}$ patch. d) At observation time, after the long-wavelength mode has reentered the horizon (solid circle indicating present horizon). The FNC patch now coincides with the usual Fermi coordinate patch, which is much larger than the small-scale modes which have been processed by nonlinear evolution since horizon entry (distorted thick wavy line).

The inverse of this transformation is at linear order

$$\begin{aligned} x^0(\bar{x}_F^\alpha) &= \bar{x}_F^0 + \frac{1}{2} \int_0^{\bar{x}_F^0} h_{00}(\tau) d\tau + (v_i + h_{0i}) \bar{x}_F^i \\ &\quad - \frac{1}{4} [h_{i,j}^0 + h_{j,i}^0 - h_{ij}'] \bar{x}_F^i \bar{x}_F^j + \mathcal{O}[(\bar{x}_F^i)^3] \\ x^k(\bar{x}_F^\alpha) &= v^k(\bar{x}_F^0 - \tau_F) + \bar{x}_F^k - \frac{1}{2} h_i^k \bar{x}_F^i \\ &\quad - \frac{1}{4} [h_{i,j}^k + h_{j,i}^k - h_{ij}'] \bar{x}_F^i \bar{x}_F^j + \mathcal{O}[(\bar{x}_F^i)^3]. \end{aligned} \quad (13)$$

The metric in $\overline{\text{FNC}}$ then becomes

$$\begin{aligned} \bar{g}_{00}^F &= -1 + \frac{1}{2} [\partial_m \partial_l h_{00} + h_{lm}'' - 2\partial_{[l} h_{0m}']] \bar{x}_F^l \bar{x}_F^m \quad (14) \\ \bar{g}_{0i}^F &= -\frac{2}{3} [\partial_{[i} \partial_l h_{0m}] + \partial_{[m} h_{li}']] \bar{x}_F^l \bar{x}_F^m \\ \bar{g}_{ij}^F &= \delta_{ij} - \frac{1}{3} [\partial_{[j} \partial_l h_{im}] + \partial_{[m} \partial_i h_{lj}']] \bar{x}_F^l \bar{x}_F^m. \end{aligned}$$

Here, $X_{(ab)} = (X_{ab} + X_{ba})/2$, $X_{[ab]} = (X_{ab} - X_{ba})/2$, and primes denote derivatives with respect to τ , while spatial derivatives are with respect to x^i . The linear and quadratic coefficients in the coordinate transform Eq. (12) are uniquely determined by the requirement that the lowest order contribution to $\bar{g}_{\mu\nu}^F - \eta_{\mu\nu}$ is order $(\bar{x}_F^i)^2$. However, given that the quadratic corrections are determined by the cubic order terms in the coordinate transformation, one might wonder whether the quadratic corrections to $\bar{g}_{\mu\nu}^F$ are unique. One can show (see App. D) that \bar{g}_{00}^F and \bar{g}_{0i}^F are indeed unique. However, there is freedom to change the quadratic term in \bar{g}_{ij}^F if one allows

the spatial coordinate lines to be non-geodesic at order $(x^i)^3$.

In summary,

$$\bar{g}_{\mu\nu}^F = \eta_{\mu\nu} + \mathcal{O}([h_{ij}'', \partial_j h_{\mu i}', \partial_i \partial_j h_{\mu\nu}] \bar{x}_F^2). \quad (15)$$

Let us disregard the corrections $\propto h_{ij}'', \partial_j h_{\mu i}'$ for the moment. The remaining corrections to the conformal Fermi frame metric scale as second derivatives of the metric perturbation multiplied by the spatial coordinate squared. Thus, instead of being order $(Hx_F)^2$ as the usual FNC corrections, they are of order $(k_L \bar{x}_F)^2$, where k_L is the comoving wavenumber of the perturbation (in slight abuse of notation, we drop the bar over k since we will only be dealing with comoving Fourier wavenumbers). Within the $\overline{\text{FNC}}$ patch of size $\sim k_L^{-1}$, we can remove the effects of the long-mode perturbation (in comoving coordinates) *at all times* — through horizon exit and reentry.

This very useful result only holds if $h_{ij}'', \partial_j h_{\mu i}'$ are of order $\partial_i \partial_j h_{\mu\nu}$ or smaller, which is the case for all models of single field inflation in which the background is an attractor solution. In principle one can engineer a model in which, for a short period of time, the background is not close to an attractor solution, but rather it is evolving towards one. In this case, the superhorizon modes can have, for a short period of time, sizable time derivatives [19, 20]. For example, if h_{ij}'' is larger than $\partial_i \partial_j h_{\mu\nu}$, corresponding to significant superhorizon evolution of the spatial metric perturbation, then we cannot follow the conformal Fermi frame through the entire duration of inflation. In case the evolving part of h_{ij} is isotropic (or

can be made so by a suitable change of coordinates),

$$h_{ij}(\mathbf{0}, \tau) = f(\tau)\delta_{ij} + \mathcal{O}(\partial_i \partial_j h_{\mu\nu}), \quad (16)$$

where $\mathbf{0}$ denotes the origin in comoving coordinates around which we construct the $\overline{\text{FNC}}$ frame, then we can absorb its effect into a modified scale factor,

$$\tilde{a}(\tau) = a(\tau) \left[1 - \frac{1}{2}f(\tau) \right]. \quad (17)$$

In a modified $\overline{\text{FNC}}$ frame constructed with the scale factor $\tilde{a}(\tau)$ the rapidly evolving part now disappears. We thus see that in this case, a long-wavelength metric perturbation does not simply shift the time and rescale spatial coordinates, but rather corresponds to a change in the entire background cosmology. This is possible only because the original “unperturbed” background cosmology, namely $f(\tau) = 0$, was unstable (i.e. not an attractor). Given that the generation of small-scale perturbations depends on the background cosmology, we in general expect a non-trivial coupling between very long wavelength metric perturbations and small-scale modes in this case. If h''_{ij} is *not* isotropic, then we cannot absorb its effect into a modified scale factor. Instead, the different spatial coordinates are rescaled differently, leading to an anisotropically expanding Universe of Bianchi I type.

Similar arguments apply to the case where $\partial_j h'_{\mu i}$ is larger than $\partial_i \partial_j h_{\mu\nu}$. Here the lowest order effect will be a Bianchi I-type Universe, since isotropy is necessarily violated.

A. Bispectra in global and $\overline{\text{FNC}}$ coordinates

Let us assume that we can calculate the statistics of a scalar field $\vartheta(\tau, \mathbf{x})$ in the global coordinates Eq. (10). Specifically, we consider fluctuations at a given epoch during inflation, typically evaluated right after horizon crossing. For simplicity, we will restrict to a scalar field, although the extension to other spins is straightforward. Given the results from Sec. II and the coordinate transform Eqs. (12)–(13), we can then immediately derive the transformation of the two-point correlation $\xi_\vartheta(\mathbf{r}, \tau)$ measured in a patch around position \mathbf{x}_0 at conformal time τ from the global coordinates to the $\overline{\text{FNC}}$ frame. Note that if ϑ is non-Gaussian, then the correlation function measured in a given patch will correlate with long-wavelength perturbations.

In the present case, the primed coordinate system of Sec. II is \bar{x}^α_F , while x^ν denote the global coordinates in the gauge chosen. We can then read off a^i_j and $\Delta\tau$ from Eq. (13):

$$a^i_j = \frac{1}{2}h^i_j \quad (18)$$

$$\Delta\tau = \frac{1}{2} \int_0^{x'^0} h_{00}(\tau) d\tau. \quad (19)$$

With this, Eq. (4) yields

$$\begin{aligned} \bar{\xi}(\bar{\mathbf{r}}; \bar{\tau}) = & \quad (20) \\ \left[1 - \frac{1}{2}h_{ij}\bar{r}^i\bar{\partial}_{\bar{r}}^j + \frac{1}{2} \left(\int_0^{\tau'} h_{00} d\tau \right) \partial_{\bar{\tau}} \right] \xi(\bar{\mathbf{r}}, \bar{\tau}), \end{aligned}$$

where $\bar{\xi}(\bar{\mathbf{r}}, \bar{\tau})$ denotes the correlation function of ϑ in the local $\overline{\text{FNC}}$ frame. Now consider the bispectrum $\langle X(\mathbf{k}_L)\vartheta(\mathbf{k}_1)\vartheta(\mathbf{k}_2) \rangle$, where X is any perturbation. The results from Sec. II immediately show how this bispectrum transforms into $\overline{\text{FNC}}$. Denoting $\overline{\text{FNC}}$ quantities as \bar{X} , $\bar{\vartheta}$, we obtain

$$\begin{aligned} B_{\bar{X}\bar{\vartheta}\bar{\vartheta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2; \bar{\tau}) & \stackrel{\text{squeezed}}{=} \\ & \left[\frac{1}{2}P_{Xh}(k_L)P_\vartheta(k_S) \frac{d \ln(k_S^3 P_\vartheta(k_S))}{d \ln k_S} \right. \\ & \quad + \frac{1}{2}P_{Xh^T_{ij}}(k_L) \frac{k_S^i k_S^j}{k_S^2} \frac{d P_\vartheta(k_S)}{d \ln k_S} \\ & \quad \left. + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\vartheta(k_S) \right]_{\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2} \\ & + B_{X\vartheta\vartheta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2; \bar{\tau}), \quad (21) \end{aligned}$$

where all correlations on the r.h.s. are evaluated at $\bar{\tau}$. Here $h = h^i_i/3$ is the trace of the spatial metric perturbation while h^T_{ij} is the trace-free part. The significance of this result will become clear in the application to single-field inflation which we will consider next.

B. Single-field inflation in comoving gauge

In the following, we restrict $h_{\mu\nu}$ to comoving gauge. In the notation of [4], we have

$$\begin{aligned} h_{00} &= -2N_1 \\ h_{0i} &= N_i \\ h_{ij} &= 2\zeta\delta_{ij} + h^T_{ij}(\tau, \mathbf{x}), \quad (22) \end{aligned}$$

where h^T_{ij} is transverse-traceless and contains the tensor perturbations. For the attractor solution of single-field inflation, the constraint equations in this gauge yield [4]

$$N_1 = \frac{\dot{\zeta}}{H} = \frac{\zeta'}{aH} \propto \frac{k^2}{(aH)^2}. \quad (23)$$

To order k_L^2/k_S^2 , we can thus neglect the contribution from the time shift $\Delta\tau = -\int N_1 d\tau$. Note that this is merely a consequence of the particular gauge chosen. The only remaining contribution to the transformation of the bispectrum [Eq. (21)] then comes from ζ and the tensors h^T_{ij} . For the scalar contribution, we obtain

$$\begin{aligned} B_{\bar{X}\bar{\vartheta}\bar{\vartheta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) & \stackrel{S}{=} P_{X\zeta}(k_L)P_\vartheta(k_S) \left[3 + \frac{\partial \ln P_\vartheta(k_S)}{\partial \ln k_S} \right] \\ & + B_{X\vartheta\vartheta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2), \quad (24) \end{aligned}$$

where throughout $\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2 = (\mathbf{k}_1 - \mathbf{k}_2)/2$. We are in particular interested in the bispectrum of the curvature perturbation ζ , i.e. $\vartheta = \zeta$. Since ζ is not a scalar, our derivation in Sec. II does not strictly apply. However, it is straightforward to show that for the purposes of this transformation, the small-scale ζ modes behave as a scalar. Recall that only the spatial transformation is relevant at the order we are interested. The comoving gauge condition, $\delta\phi = 0$ where $\delta\phi$ is the inflaton perturbation is thus still satisfied in the FNC frame. Further, the spatial components of the metric transform as

$$\bar{g}_{ij}^F(\bar{x}_F) = \frac{\partial x^k}{\partial \bar{x}_F^i} \frac{\partial x^l}{\partial \bar{x}_F^j} g_{kl}(x). \quad (25)$$

We now write $\zeta(x) = \zeta_L(x) + \zeta_S(x)$, separating into long- and short-wavelength pieces on the scale of the patch within which the correlation function is measured. Then, the transformation to FNC removes ζ_L up to second derivatives, while ζ_S is not affected since it does not contain any long-wavelength components. We thus obtain

$$\begin{aligned} \bar{\zeta}(\bar{x}_F) &= \zeta_S(x) + \mathcal{O}(\partial_i \partial_j \zeta_L) \\ \bar{\zeta}_S(\bar{x}_F) &= \zeta_S(x). \end{aligned} \quad (26)$$

Thus, the short-wavelength ζ perturbations transform effectively as a scalar [Eq. (1)]. We obtain for the bispectrum of curvature perturbations in the FNC frame:

$$\begin{aligned} B_{\bar{\zeta}\bar{\zeta}\bar{\zeta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) &= P_\zeta(k_L) P_\zeta(k_S) \frac{\partial \ln(k_S^3 P_\zeta(k_S))}{\partial \ln k_S} \\ &+ B_{\zeta\zeta\zeta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2), \end{aligned} \quad (27)$$

where $B_{\zeta\zeta\zeta}$ is the three-point function calculated in comoving gauge. As shown in [4, 21], this is in the squeezed limit given by

$$B_{\zeta\zeta\zeta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) = -[n_s - 1] P_\zeta(k_L) P_\zeta(k_S), \quad (28)$$

where n_s is defined through $P_\zeta(k) \propto k^{-4+n_s}$. Eq. (28) is usually referred to as the *consistency relation*. We now see that the first term in Eq. (27) exactly cancels the contribution from $B_{\zeta\zeta\zeta}$, leading to

$$B_{\bar{\zeta}\bar{\zeta}\bar{\zeta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) = \mathcal{O}\left(\frac{k_L}{k_S}\right)^2. \quad (29)$$

We thus conclude that *in single-field inflation, the bispectrum in the squeezed limit is zero in the conformal Fermi frame, with corrections going as $(k_L/k_S)^2$.*

The case for non-scalar metric perturbations follows analogously. Decomposing the long-wavelength metric perturbation into polarization states,

$$h_{ij}^T(\mathbf{k}_L) = \sum_{s=+, \times} e_{ij}^s(\hat{\mathbf{k}}_L) h_s^T(\mathbf{k}_L), \quad (30)$$

we obtain

$$\begin{aligned} B_{\bar{X}\bar{\vartheta}\bar{\vartheta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) &\stackrel{T}{=} \\ &\frac{1}{2} \sum_s P_{Xh_s}(k_L) P_\vartheta(k_S) e_{ij}^s(\hat{\mathbf{k}}_L) \hat{k}_S^i \hat{k}_S^j \frac{\partial \ln P_\vartheta(k_S)}{\partial \ln k_S} \\ &+ B_{X\vartheta\vartheta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \end{aligned} \quad (31)$$

Assuming that the different polarization states are statistically independent, we obtain for the tensor-scalar-scalar bispectrum [again using the transformation property Eq. (26)]

$$\begin{aligned} B_{\bar{h}_s\bar{\zeta}\bar{\zeta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) &= \\ &\frac{1}{2} P_{h_s}(k_L) P_\zeta(k_S) e_{ij}^s(\hat{\mathbf{k}}_L) \hat{k}_S^i \hat{k}_S^j \frac{\partial \ln P_\zeta(k_S)}{\partial \ln k_S} \\ &+ B_{h_s\zeta\zeta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \end{aligned} \quad (32)$$

The squeezed-limit bispectrum in comoving gauge was also derived in [4]:

$$\begin{aligned} B_{h_s\zeta\zeta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) &= -P_{h_s}(k_L) e_{ij}^s(\hat{\mathbf{k}}_L) \hat{k}_S^i \hat{k}_S^j \frac{d}{dk_S^2} P_\zeta(k_S) \\ &= -P_{h_s}(k_L) e_{ij}^s(\hat{\mathbf{k}}_L) \hat{k}_S^i \hat{k}_S^j \frac{1}{2} \frac{d}{d \ln k_S} P_\zeta(k_S). \end{aligned} \quad (33)$$

We again see that the two terms in Eq. (32) cancel in single-field inflationary models. Thus, *in single-field inflation, the tensor-scalar-scalar bispectrum vanishes in the squeezed limit in FNC, so that there are no correlations between long-wavelength tensor modes and small-scale fluctuations in this frame.* The lowest correction are again of order $(k_L/k_S)^2$.

We can phrase the main result of this section as follows: at leading order, the squeezed-limit three-point correlations in single-field inflation, which obey the “consistency relation”, are equivalent to the statement that there is no correlation between infinitely long and short wavelength modes in the conformal Fermi frame, specifically,

$$B_{\bar{h}\bar{\zeta}\bar{\zeta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) = \mathcal{O}\left(\frac{k_L}{k_S}\right)^2, \quad (34)$$

where h stands for any component of h_{ij} , and $k_S = |\mathbf{k}_1 - \mathbf{k}_2|/2$.

Since this latter is a physical, gauge-invariant statement, it is expected to hold not only at leading order in spacetime perturbations but at higher orders as well. The FNC approach can also help elucidate why there is no such correlation up to order k_L^2 in single field models. One might wonder why such a correlation cannot be imprinted at early times when the long wavelength perturbation was inside the horizon, $k_L \gg aH$. Far inside the horizon, we can neglect gravity and are essentially dealing with a scalar field in vacuum which adiabatically tracks the slowly evolving background, and a given mode is only excited once its wavelength becomes of order the horizon. Contributions to correlations from within the horizon are exponentially suppressed [9, 10] (see also [22]). On the

other hand, when $k_S \sim aH$, the long-wavelength mode is far outside the horizon and its effects can be removed by a coordinate transformation up to order k_L^2 .

IV. CONNECTION TO LATE-TIME OBSERVATIONS

As explained in Sec. III, we can follow the $\overline{\text{FNC}}$ patch all the way through the end of inflation and horizon re-entry of both long and short modes, provided that perturbations do not evolve significantly when they are outside the horizon. In this section, we show how the squeezed-limit correlations in the $\overline{\text{FNC}}$ frame can be related to observations made from Earth today.

Let us assume we observe correlations of some field $\delta(\mathbf{x})$ at late times, i.e. at or after recombination. In linear theory, assuming adiabatic initial conditions, we can relate δ to the curvature perturbation through some transfer function α ,

$$\delta(\mathbf{k}, \tau) = \alpha(k, \tau)\zeta(\mathbf{k}). \quad (35)$$

For matter density perturbations, α is defined in Eq. (71) below. Similarly, we assume the long-wavelength perturbation is evolved with some transfer function,

$$h_{ij}(\mathbf{k}_L, \tau) = G(\tau)h_{ij}(\mathbf{k}_L), \quad (36)$$

where we have assumed the $k_L \rightarrow 0$ limit (the lowest order k_L dependence of $G(\tau)$ will be k_L^2 ; of course, G will differ for scalar and tensor perturbations). As long as superhorizon perturbations evolve slowly (in the sense that $h'', \partial_j h'$ are smaller than or of order $\partial_i \partial_j h$), the conformal Fermi coordinate patch, i.e. the region over which corrections to the $\overline{\text{FNC}}$ metric are small, is essentially constant in (comoving) size throughout horizon exit and reentry of the short-wavelength modes. As we have seen, this applies in particular to single-field inflation models. Thus, at the conformal time τ_{em} at which the photons we observe today were emitted, the metric in $\overline{\text{FNC}}$ is

$$\bar{g}_{\mu\nu}^F = a^2(\tau_{\text{em}})\eta_{\mu\nu} + \mathcal{O}(\partial_i \partial_j h). \quad (37)$$

If we transform coordinates through

$$\begin{aligned} x_F^0 &= \int^\tau a(\tau') d\tau' \equiv t(\tau) \\ x_F^i &= a(\tau)x_F^i, \end{aligned} \quad (38)$$

the metric in the coordinates $\{x_F^\mu\}$ becomes

$$g_{\mu\nu}^F = \eta_{\mu\nu} + \mathcal{O}[(Hx_F^i)^2, (\partial\partial h_{\mu\nu})x_F^2]. \quad (39)$$

In other words, x_F^μ are the usual Fermi coordinates (FNC) defined around the same timelike geodesic as the $\overline{\text{FNC}}$. Depending on whether the long-wavelength modes for which we constructed the $\overline{\text{FNC}}$ patch have entered the horizon, either the order H or the order $\partial\partial h_{\mu\nu}$ corrections will be dominant; however, this is not relevant

for the discussion that follows. Eq. (38) corresponds to a rescaling of the time coordinate, leaving the timelike unit vector and $\tau = \text{const}$ hypersurfaces unchanged, and a time-dependent rescaling of the spatial coordinates. Since the spatial rescaling is the same everywhere on a $x_F^0 = \text{const}$ hypersurface, this implies that, for models that obey the consistency condition, Eq. (34) is still valid at late times in the FNC frame:

$$B_{h_F \delta_F \delta_F}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2; \tau_{\text{em}}) = \mathcal{O}\left(\frac{k_L}{k_S}\right)^2, \quad (40)$$

where h_F, δ_F denote perturbations in the FNC frame. Recall that the squeezed limit of three-point functions corresponds to the modulation of local two-point functions by long-wavelength perturbations. Thus, another way of phrasing this result is that a surface of constant correlation in the FNC frame, $\xi_{\delta_F}(r_F) = \text{const}$, defines a standard ruler—a fixed spatial scale—as considered in [6], with corrections proportional to second derivatives of $h_{\mu\nu}$ only. Hence, in standard single-field inflation and any other case where Eq. (34) holds, the ruler scale r_F is statistically the same everywhere on a $x_F^0 = \text{const}$ hypersurface, and there is no correlation with long-wavelength perturbations.

The *apparent* correlations induced between long-wavelength modes and small-scale correlations are then given by the ruler perturbations derived in [6]. We can use their results together with Sec. II to derive these contributions to observed squeezed-limit three-point functions, and make the connection with known results.

In general, there are two effects modifying the observed two-point correlation within a given patch. First, there is the transformation from the local FNC x_F^μ to the observed comoving coordinates \tilde{x}^μ , which are inferred from the observed position $\hat{\mathbf{n}}$ of the source in the sky and its redshift \tilde{z} through $\tilde{x}^0 = \tau_0 - \bar{\chi}(\tilde{z})$, $\tilde{x}^i = \hat{n}^i \bar{\chi}(\tilde{z})$, where $\bar{\chi}$ is the comoving distance-redshift relation in the background (in case of the CMB, this is slightly modified, as we will discuss in Sec. IV A). Let us assume we observe a scalar field $\tilde{\rho}(\tilde{x})$ (whose perturbation is δ). Then, as described in Sec. II, $\tilde{\rho}$ is given in terms of the field in the Fermi frame ρ as

$$\tilde{\rho}(\tilde{x}) = \rho(x_F(\tilde{x})). \quad (41)$$

At fixed observed redshift, we can write the transformation from x_F to \tilde{x} as

$$\begin{aligned} \tilde{x}^0 &= x_F^0 + \frac{1}{\tilde{a}\tilde{H}}\mathcal{T} \\ \tilde{x}^i &= \frac{1}{\tilde{a}}A^i_j x_F^j. \end{aligned} \quad (42)$$

Here $\tilde{a} = (1 + \tilde{z})^{-1}$, \mathcal{T} is the perturbation in proper time from a constant observed redshift surface [15], and A^i_j can be seen as the generalization to three dimensions of the magnification matrix. Note that \mathcal{T} and A^i_j are gauge-invariant. Specifically, using the notation of [6],

A_{ij} is given by

$$\begin{aligned} A_{ij} &= \delta_{ij} + a_{ij} \\ a_{ij} &= \mathcal{C}\hat{n}_i\hat{n}_j + \hat{n}_{(i}\mathcal{P}_{j)k}\mathcal{B}^k + \mathcal{P}_{ik}\mathcal{P}_{jl}\mathcal{A}^{kl}, \end{aligned} \quad (43)$$

where $\mathcal{P}^{ij} = \delta^{ij} - \hat{n}^i\hat{n}^j$ is the projection operator perpendicular to the line of sight. \mathcal{C} , \mathcal{B}_i , and \mathcal{A}_{ij} are the gauge-invariant ruler perturbations derived in terms of the metric perturbations in [6]. For example, the transverse matrix \mathcal{A}_{ij} contains the magnification and shear.

The second effect is a rescaling of the observed field by projection effects. If there are multiplicative projection effects, then we have

$$\rho_{\text{obs}}(\tilde{x}) = [1 + c(\tilde{x})] \tilde{\rho}(\tilde{x}) = [1 + c(\tilde{x})] \rho(x_F(\tilde{x})). \quad (44)$$

Here we think of c as averaged over the patch within which we measure the correlation function of small-scale perturbations. For example, in case ρ is the number density of some tracer, then projection effects such as gravitational lensing modify the physical volume that corresponds to a fixed region in the observed coordinates \tilde{x} , thus rescaling the number density. Gravitational redshift and ISW effect lead to an analogous rescaling in case of the CMB. The factor $1 + c$ also rescales the fluctuations in ρ within the region considered, and correspondingly rescales the correlation function by $(1 + c)^2$, leading to

$$\begin{aligned} \tilde{\xi}(\tilde{\mathbf{r}}, \tilde{\tau}) &= [1 + 2c(\tilde{x})] \xi(\mathbf{r}_F, \tau_F) \\ &= \left[1 - a_{ij}(\tilde{x}) \tilde{r}^i \partial_{\tilde{r}}^j + \frac{1}{\tilde{a}\tilde{H}} \mathcal{T}(\tilde{x}) \partial_{\tilde{\tau}} + 2c(\tilde{x}) \right] \xi(\tilde{\mathbf{r}}; \tilde{\tau}) \end{aligned} \quad (45)$$

where we have used Eq. (4) in the second line. Straightforward application of the results of Sec. II then yields the bispectrum of the observed density perturbation $\tilde{\delta}$ in the squeezed limit, in terms of the projection effects and the bispectrum of δ_F in the FNC frame:

$$\begin{aligned} B_{\tilde{\delta}\tilde{\delta}\tilde{\delta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) &= B_{\delta_F\delta_F\delta_F}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) \\ &+ 2P_{\delta_F c}(k_L)P_{\delta_F}(k_S) + \frac{1}{\tilde{a}\tilde{H}}P_{\delta_F\mathcal{T}}(k_L)\frac{\partial}{\partial\tau}P_{\delta_F}(k_S) \\ &+ \frac{1}{3}\frac{\partial\ln(k_S^3P_{\delta_F}(k_S))}{\partial\ln k_S}P_{\delta_F a}(k_L)P_{\delta_F}(k_S) \\ &+ P_{\delta_F a_{ij}^T}(k_L)\hat{k}_S^i\hat{k}_S^j\frac{\partial P_{\delta_F}(k_S)}{\partial\ln k_S}, \end{aligned} \quad (46)$$

where $\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2$. The bispectrum of a two-dimensional projected field (in the flat-sky limit) correspondingly becomes

$$\begin{aligned} B_{\tilde{\delta}\tilde{\delta}\tilde{\delta}}(\ell_L, \ell_1, \ell_2) &= B_{\delta_F\delta_F\delta_F}(\ell_L, \ell_1, \ell_2) \\ &+ 2C_{\delta_F c}(\ell_L)C_{\delta_F}(\ell_S) + \frac{1}{\tilde{a}\tilde{H}}C_{\delta_F\mathcal{T}}(\ell_L)\frac{\partial}{\partial\tau}C_{\delta_F}(\ell_S) \\ &+ \frac{1}{2}\frac{\partial\ln(\ell_S^2P_{\delta_F}(\ell_S))}{\partial\ln\ell_S}C_{\delta_F a}(\ell_L)C_{\delta_F}(\ell_S) \\ &+ C_{\delta_F a_{ij}^T}(\ell_L)\hat{\ell}_S^i\hat{\ell}_S^j\frac{\partial C_{\delta_F}(\ell_S)}{\partial\ln\ell_S}, \end{aligned} \quad (47)$$

where again $\ell_S = \ell_1 + \ell_L/2$.

The bispectrum $B_{\delta_F\delta_F\delta_F}$ in FNC frame is equivalent up to transfer functions (and radial projection, in the 2D case) to the bispectrum in $\overline{\text{FNC}}$, which as we have seen vanishes in the squeezed limit for single-field inflation. In this case, the terms due to “projection effects” are the only remaining contributions.

In the following we present two applications of Eqs. (46)–(47): the CMB bispectrum in the squeezed limit, and the scale-dependent non-Gaussian halo bias.

A. CMB bispectrum

In this section, we illustrate the main result of the previous section on the observed squeezed-limit bispectrum, Eq. (47), with the CMB. We will make direct connection with the results of [13]. A more sophisticated and accurate treatment which is based on an closely related approach has been presented in [14].

We assume single-field inflation so that the $\overline{\text{FNC}}$ frame contribution vanishes. We will adopt the conformal-Newtonian gauge in this section,

$$ds^2 = a^2(\tau)[-(1 + 2\Phi)d\tau^2 + (1 - 2\Psi)d\mathbf{x}^2], \quad (48)$$

although it is straightforward to derive the results in a general gauge.

The observed CMB photons originate from the last scattering surface, which occurred at a fixed physical age t_* of the Universe, that is, at constant proper time $t_F = t_*$ for the comoving primordial plasma. The proper time of a comoving source passing through \mathbf{x} at coordinate time τ is given in the metric convention Eq. (48) by

$$t_F|_{\tau, \mathbf{x}} = \bar{t}(\tau) + \int_0^\tau \Phi(\mathbf{x}, \tau')a(\tau')d\tau', \quad (49)$$

where $\bar{t}(\tau)$ and $\bar{\tau}(t)$ are the physical-time – conformal time relations in the background.

The value of t_* is obtained by combining atomic physics with the mean observed temperature of the CMB today. The CMB temperature perturbations on scales that were super-horizon at recombination ($\ell \lesssim 100$) originate entirely from projection effects; in other words, the large-scale CMB temperature perturbations can be seen as a special case of the evolving ruler described in [15]. Essentially, the standard ruler is in this case given by the photon occupation number I_ν/ν^3 . In fact, since lensing conserves surface brightness, the only contribution to the fractional CMB temperature perturbation $\Theta(\hat{\mathbf{n}})$ is the redshift perturbation on a fixed proper time surface, which is equivalent to minus the quantity \mathcal{T} defined in [15]:

$$\Theta(\hat{\mathbf{n}}) \equiv \frac{T(\hat{\mathbf{n}})}{\bar{T}} - 1 = \frac{d\ln T(a)}{d\ln a}\mathcal{T} = -\mathcal{T} \quad (50)$$

where we have used that $T \propto a^{-1}$ for a free-streaming blackbody. As shown in [15], this immediately yields the CMB

temperature perturbation in the gauge Eq. (48) as

$$\begin{aligned}\Theta(\hat{\mathbf{n}}) &= -\tilde{H}_* \int_0^{t_*} \Phi[\mathbf{x}, \bar{\tau}(t')] dt' + H_0 \int_0^{t_0} \Phi(\mathbf{0}, \bar{\tau}(t)) dt \\ &\quad - \Phi_o + \Phi - v_{\parallel} + v_{\parallel o} + \delta\nu_{\text{ISW}} \\ &\simeq \frac{1}{3}\Phi - v_{\parallel} + H_0 \int_0^{t_0} \Phi(\mathbf{0}, \bar{\tau}(t)) dt - \Phi_o + v_{\parallel o} + \delta\nu_{\text{ISW}}\end{aligned}\quad (51)$$

$$\delta\nu_{\text{ISW}} = \int_0^{\chi_*} d\chi [\Phi' + \Psi'] . \quad (52)$$

In the second line, we have assumed that the relevant perturbations are superhorizon, $\Phi(\mathbf{x}) \approx \text{const}$ (since we are interested in $\ell \lesssim 100$ where the acoustic contributions can be neglected), and that recombination happened long after matter-radiation equality so that $H_* t_* = 2/3$.

1. Transforming from FNC to observer coordinates

As we have seen, the CMB temperature on large scales is modified by the temperature perturbation Eq. (51), which is entirely due to the effects on photons as they propagate from the last scattering surface to the observer. Thus, the CMB temperature that would be measured in the local Fermi frame at emission is rescaled by $1 + \Theta(\hat{\mathbf{n}})$ in the observer frame, so that in the notation of Sec. IV, $c(\hat{\mathbf{n}}) = \Theta(\hat{\mathbf{n}})$.

The second ingredient is the coordinate transformation from FNC to the observer frame. Projected onto the sky, the matrix a_{ij} becomes the usual weak lensing distortion tensor \mathcal{A}_{ij} , which we can write to linear order as

$$a_{ij} = \begin{pmatrix} \mathcal{M}/2 + \gamma_1 & \gamma_2 \\ \gamma_2 & \mathcal{M}/2 - \gamma_1 \end{pmatrix}. \quad (53)$$

In particular, $a = \mathcal{M}$. The magnification (a gauge-invariant quantity) is easily adapted from the results of [6]. The key difference is the meaning of the perturbation to the logarithm of the scale factor at emission, which here we call $\Delta \ln a_*$. By definition,

$$\Delta \ln a_* \equiv \frac{a(x_{\text{em}}^0)}{a_*} - 1. \quad (54)$$

Whereas the quantity $\Delta \ln a$ of [6] is derived for photons arriving with a fixed observed redshift, in the present case we need the corresponding expression for a constant proper time at emission. This is easily derived from the expression for the proper time Eq. (49). Requiring $t_F = t_*$ and solving for τ yields

$$\begin{aligned}a(x_{\text{em}}^0) &= a \left[\bar{\tau} \left(t_* - \int_0^{\tau_*} \Phi(\mathbf{x}, \tau) a d\tau \right) \right] \\ &= a_* \times \left[1 - H_* \int_0^{\tau_*} \Phi(\mathbf{x}, \tau) a d\tau \right],\end{aligned}\quad (55)$$

which leads to

$$\frac{a(x_{\text{em}}^0)}{a_*} = 1 - H_* \int_0^{\tau_*} \Phi(x, \tau) a d\tau, \quad (56)$$

so that

$$\Delta \ln a_* = -H_* \int_0^{\tau_*} \Phi(x, \tau) a d\tau. \quad (57)$$

We now use this result in the general-gauge expression Eq. (51) of [6],

$$\mathcal{M} = -2\Delta \ln a_* - \frac{1}{2} (h_i^i - h_{\parallel}) + 2\hat{\kappa} - \frac{2}{\chi_*} \Delta x_{\parallel}. \quad (58)$$

In Eq. (58), $\hat{\kappa}$ is the coordinate convergence, i.e. the transverse (with respect to the line of sight) divergence of the transverse displacements, and Δx_{\parallel} is the displacement along the line of sight. In the metric Eq. (48) we have [6]

$$\hat{\kappa} = -v_{\parallel o} + \frac{1}{2} \int_0^{\chi_*} d\chi \frac{\chi}{\chi_*} (\chi_* - \chi) \nabla_{\perp}^2 (\Phi + \Psi) \quad (59)$$

$$\begin{aligned}\Delta x_{\parallel} &= \int_0^{\chi_*} d\chi [\Phi + \Psi] - \frac{1}{a_* H_*} \Delta \ln a_* \\ &\quad - \int_0^{t_0} \Phi(\mathbf{0}, t) dt\end{aligned}\quad (60)$$

$$\Delta \ln a_* = -\frac{2}{3} \Phi, \quad (61)$$

where $H_* = H(a_*)$, and we have made the same assumptions about superhorizon perturbations and matter domination as in Eq. (51). The magnification then becomes

$$\begin{aligned}\mathcal{M} &= \frac{4}{3} \Phi - (-2\Psi) + 2\hat{\kappa} - \frac{2}{\chi_*} \Delta x_{\parallel} \\ &= \frac{10}{3} \Phi + 2\hat{\kappa} - \frac{2}{\chi_*} \Delta x_{\parallel}.\end{aligned}\quad (62)$$

Here we have used that $\Psi = \Phi$ during matter domination.

Using the relation between convergence κ and shear γ in ℓ -space, we then have in the flat-sky approximation

$$\begin{aligned}C_{Tc}(\ell) &= C_{TT}(\ell) \\ C_{Ta}(\ell) &= C_{T\mathcal{M}}(\ell) = \frac{10}{3} C_{T\Phi_{\text{em}}}(\ell) + 2C_{T\hat{\kappa}}(\ell) - \frac{2}{\chi_*} C_{T\Delta x_{\parallel}} \\ &\approx 10C_{TT}^{\text{no ISW}}(\ell) + 2C_{T\hat{\kappa}}(\ell)\end{aligned}\quad (63)$$

$$C_{Th_{ij}^T}(\ell) = C_{T\hat{\kappa}}(\ell) \left[2\frac{\ell_i \ell_j}{\ell^2} - \delta_{ij} \right]. \quad (64)$$

In Eq. (63), we have approximated $C_{T\Phi_{\text{em}}}(\ell)/3$ as the CMB temperature power spectrum without ISW contribution, and neglected the contribution from the line-of-sight displacement Δx_{\parallel} . The latter is small due to cancellation along the line of sight except for the very smallest ℓ for the time delay [first term in Eq. (60)], and suppression by $(H_* \chi_*)^{-1} \sim \sqrt{1089}$ in case of the second

term of Eq. (60) (the third term in Eq. (60) only contributes to the monopole). See [14] for a quantitative evaluation of these contributions. Finally, for the contribution from the time shift, $C_{TT}(\ell_L) = -C_{TT}(\ell_L)$ since $\Theta = -\mathcal{T}$. However, the CMB power spectrum observed today $C_{TT}(\ell_S)$ only evolves on the Hubble scale today, so that

$$\frac{1}{\tilde{a}\tilde{H}}C_{TT}(\ell_L)\frac{\partial}{\partial\tau}C_{TT}(\ell_S) \sim \frac{H_0}{\tilde{a}\tilde{H}}C_{TT}(\ell_L)C_{TT}(\ell_S). \quad (65)$$

Thus, this contribution is suppressed by $\sim 1/\sqrt{1089}$ with respect to the leading contributions, and we will neglect it in what follows.

2. Squeezed-limit CMB bispectrum

Inserting these results into Eq. (47) then yields

$$\begin{aligned} B(\ell_L, \ell_1, \ell_2) &= 2C_{Tc}(\ell_L)C_{TT}(\ell_S) \\ &+ C_{Ta}(\ell_L)C_{TT}(\ell_S)\frac{1}{2}\frac{\partial \ln \ell_S^2 C_{TT}(\ell_S)}{\partial \ln \ell_S} \\ &+ C_{Ta_{ij}^T}(\ell_L)\hat{\ell}_S^i\hat{\ell}_S^j\frac{\partial C_{TT}(\ell_S)}{\partial \ln \ell_S} \\ &= \left\{ 2C_{TT}(\ell_L) + (5C_{TT}^{\text{no ISW}}(\ell_L) + C_{T\hat{\kappa}}(\ell_L))\frac{\partial \ln \ell_S^2 C_{TT}(\ell_S)}{\partial \ln \ell_S} \right. \\ &\quad \left. + \cos 2\theta C_{T\hat{\kappa}}(\ell_L)\frac{\partial \ln C_{TT}(\ell_S)}{\partial \ln \ell_S} \right\} C_{TT}(\ell_S), \quad (66) \end{aligned}$$

where $\ell_S = \ell_L + \ell_1/2$, and $\cos \theta = \hat{\ell}_L \cdot \hat{\ell}_S$. Apart from the very small contribution from Δx_{\parallel} , this is an exact result for the squeezed-limit CMB bispectrum as long as the approximation of the Sachs-Wolfe limit is accurate for modes of wavenumber ℓ_L ($\ell_L \lesssim 100$). Note that no such assumption has been made about the short-wavelength modes ℓ_S . The cross-correlation between CMB temperature and $\hat{\kappa}$ can further be decomposed as

$$C_{T\hat{\kappa}}(\ell_L) = \frac{1}{3}C_{\Phi_{\text{em}}\hat{\kappa}}(\ell_L) + C_{\delta\nu_{\text{ISW}}\hat{\kappa}}(\ell_L), \quad (67)$$

i.e. an early-time correlation with the potential Φ_{em} , and a late-time correlation between $\hat{\kappa}$ and the ISW contribution. The latter contribution has recently been detected by the Planck satellite [23].

In order to compare with Creminelli et al. [13], we neglect the ISW-lensing correlation, and use the result of [12] for early-time correlations in the Sachs-Wolfe regime,

$$\frac{1}{3}C_{\Phi_{\text{em}}\hat{\kappa}}(\ell_L) = -6 C_{TT}^{\text{no ISW}}(\ell_L). \quad (68)$$

Further, we neglect the distinction between $C_{TT}(\ell_L)$ and

$C_{TT}^{\text{no ISW}}(\ell_L)$. We then have

$$\begin{aligned} B(\ell_L, \ell_1, \ell_2) &= \left[2C_{TT}(\ell_L) - C_{TT}(\ell_L)\frac{\partial \ln \ell_S^2 C_{TT}(\ell_S)}{\partial \ln \ell_S} \right. \\ &\quad \left. - 6 \cos 2\theta C_{TT}(\ell_L)\frac{\partial \ln C_{TT}(\ell_S)}{\partial \ln \ell_S} \right] C_{TT}(\ell_S) \\ &= C_{TT}(\ell_L)C_{TT}(\ell_S) (1 + 6 \cos 2\theta) \left[2 - \frac{\partial \ln \ell_S^2 C_{TT}(\ell_S)}{\partial \ln \ell_S} \right]. \quad (69) \end{aligned}$$

This agrees with the final result of [13].

One can show (App. C of [6]) that the contribution of metric perturbations with wavenumber k to the CMB temperature as well as the ruler perturbations \mathcal{M} , γ scale as $(k/H_0)^2$ in the limit $k \rightarrow 0$. The contributions in the low- k limit can be interpreted as the lowest order corrections to our local conformal Fermi patch. Unless there is some physical coupling between currently superhorizon and subhorizon modes [24], the observable imprint of *any* superhorizon perturbation (scalar or tensor) is suppressed by $(k/H_0)^2$.

B. Non-Gaussian halo bias

For Gaussian initial conditions, the distribution of large-scale structure tracers (such as galaxies, clusters, etc.) in the large-scale limit follows the distribution of matter. More accurately, this holds on a constant proper time slice [7, 25]. While the abundance of tracers depends on the amplitude and shape of local small-scale fluctuations, in the Gaussian case these are statistically the same everywhere. That is, at fixed proper time in their respective FNC frame, all observers see statistically the same small-scale fluctuations. These then do not contribute to correlations in the tracer abundance on large scales. Non-Gaussianity in the primordial perturbations can however couple small-scale fluctuations to large-scale perturbations. As we discussed in Sec. II, this in fact precisely corresponds to the squeezed limit of (at lowest order) the three-point function. Thus, neglecting for the time being any projection effects in going from the FNC frame to the observed positions and redshifts of tracers, there is a non-Gaussian scale-dependent bias if and only if the amplitude of small-scale fluctuations in the FNC frame correlates with long-wavelength perturbations—that is, if there is a non-zero squeezed-limit bispectrum $B_{\delta_F\delta_F\delta_F}$, where δ_F denotes the matter density perturbation in the FNC frame.

Using Eqs. (35)–(36), the bispectrum of δ_F is related to the bispectrum of curvature perturbations in the $\overline{\text{FNC}}$ frame through

$$\begin{aligned} B_{\delta_F\delta_F\delta_F}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2; \tau_{\text{em}}) &= \alpha(k_L, \tau_{\text{em}})\alpha(k_1, \tau_{\text{em}})\alpha(k_2, \tau_{\text{em}}) \\ &\quad \times B_{\bar{\zeta}\bar{\zeta}\bar{\zeta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2), \quad (70) \end{aligned}$$

where

$$\alpha(k) = \frac{5}{3} \frac{2}{3} \frac{k^2 T(k) g(z)}{\Omega_m H_0^2 (1+z)} \quad (71)$$

is the relation in Fourier space between the density and the curvature perturbation, $T(k)$ is the matter transfer function normalized to unity as $k \rightarrow 0$, and $g(z)$ is the linear growth rate of the gravitational potential normalized to unity during the matter dominated epoch.

In order to derive the scale-dependent bias, we now assume that the tracer abundance is mostly sensitive to the variance of the density field on a scale R , typically chosen to correspond to the Lagrangian scale of the tracer. See [25, 26] for a more general and detailed discussion. The contribution to the tracer two-point function is then proportional to $\langle \delta_F(\mathbf{0})(\delta_{F,R})^2(\mathbf{r}) \rangle$, where $\delta_{F,R}$ denotes the FNC-frame density field smoothed on scale R . As shown in, e.g. [27, 28], the large-scale scale-dependent bias is then given by

$$\begin{aligned} \Delta b(k_L) &= \frac{\partial \ln \bar{n}_h}{\partial \sigma_R^2} \alpha^{-1}(k_L) \Delta \sigma_R^2(k_L) \\ \Delta \sigma_R^2(k_L) &= \int \frac{d^3 k_1}{(2\pi)^3} \alpha(k_1) \alpha(|\mathbf{k} + \mathbf{k}_1|) \frac{B_{\zeta\bar{\zeta}\bar{\zeta}}(k_L, k_1, |\mathbf{k} + \mathbf{k}_1|)}{P_{\zeta}(k_L)} \\ &\quad \times \tilde{W}_R(k_1) \tilde{W}_R(|\mathbf{k} + \mathbf{k}_1|), \end{aligned} \quad (72)$$

where $\tilde{W}_R(k)$ is the filter function in Fourier space, and we have dropped the τ_{em} arguments for brevity. As long as k_L is much less than the typical values of k_1 in the integrand ($k_1 \gg 0.01 h \text{ Mpc}^{-1}$), the bispectrum here is evaluated in the squeezed limit [29]. Let us perform a leading order expansion in this limit,

$$B_{\zeta\bar{\zeta}\bar{\zeta}}(k_L, k_1, |\mathbf{k} + \mathbf{k}_1|) = \left(f_0 + f_2 \frac{k_L^2}{k_S^2} \right) P_{\zeta}(k_L) P_{\zeta}(k_S),$$

where k_S is as defined before. Here, we have assumed a scale-invariant bispectrum (otherwise, the coefficients f_0, f_2 are in general functions of k_S), and no dependence on the angle between k_L and k_S as is typically the case (but see [30] for a counterexample). The lowest order piece $\propto f_0$ leads to a scale dependence of

$$\Delta b(k_L) \propto f_0 \alpha^{-1}(k_L) \propto f_0 k_L^{-2} T^{-1}(k_L), \quad (73)$$

which corresponds to the usual scale-dependent bias from local non-Gaussianity. The second part leads to a scale-dependence of

$$\Delta b(k_L) \propto f_2 k_L^2 \alpha^{-1}(k_L) \propto f_2 T^{-1}(k_L), \quad (74)$$

which is much weaker and only relevant on scales $k_L \gtrsim 0.02 h \text{ Mpc}^{-1}$ where the transfer function departs from unity. Since $f_0 = 0$ in single field inflation [Eq. (34)], we conclude there is *no scale-dependent bias induced on scales* $k_L \lesssim 0.02 h \text{ Mpc}^{-1}$ in these models.

The scale-dependent bias discussed here refers to the FNC frame of the tracers; there are contributions to the

observed scale-dependent bias from projection effects, i.e. from transforming from the tracer FNC frame to observed positions and redshifts (in our own FNC frame). Those have been derived in [7, 31–34], but are generally quite small. Matching to the scale-dependent bias from local non-Gaussianity, they correspond to $f_{\text{NL}} < 1$ for a wide range of tracer parameters [34].

V. CONCLUSIONS

In this work we have presented a simple and complete framework to translate the squeezed limit of any primordial three-point function into observables such as the squeezed limit of the CMB temperature bispectrum and the scale-dependent halo bias. In Fig. 1 we have spelled out the various steps of the computation and the respective choices of gauge. The different gauges have been chosen in such a way that the whole computation from horizon exit during inflation until observation on earth can be performed using only linear perturbation theory. This required the introduction of a conformal version of the well known Fermi Normal Coordinates (used in [9, 10] for inflationary correlators), in which the spacetime is locally FLRW as opposed to Minkowski. As an example we have applied our formalism to standard single-field slow-roll inflation. We have shown that Maldacena’s consistency condition [4] in the squeezed limit of the scalar bispectrum implies that the signal in the CMB bispectrum and halo bias vanishes exactly. Although this result was already known, our approach provides a simple, concise and physically clear derivation. This calculation can straightforwardly be generalized to higher N-point functions in the squeezed limit.

In addition, our approach shows that, for exactly the same reason, the consistency relation for the scalar-scalar-tensor bispectrum implies the absence of any observable correlation between long-wavelength tensor modes and short-wavelength scalar perturbations (“fossils”) as proposed for example in [16] (see also [35, 36]). Taking the expression for the tensor-scalar-scalar bispectrum in comoving gauge Eq. (33) at face value, Masui and Pen [16] showed that the signal-to-noise ratio squared for detecting a scale-invariant spectrum of tensor modes with wavenumbers in the range $[k_{\text{min}}, k_{\text{max}}]$ scales as k_{min}^{-3} (if k_{min} is horizon scale or smaller). On the other hand, we have shown that in single-field inflation the only contributions to the observed correlations between long-wavelength tensor modes and small-scale perturbations are due to projection effects, which scale as $(k/aH)^2$ times the tensor perturbation where aH is to be evaluated at the time of emission [37, 38]. In this case, the signal-to-noise ratio squared roughly scales as k_{max} , and is thus dominated by the smallest observable wavelengths rather than the largest. Note that due to the oscillatory nature of the tensor mode transfer function (i.e. the fact that tensor modes propagate), the observable effects of tensor modes with $k \gg aH$ are usually highly suppressed

[38, 39]. On the other hand, just as for the scalar bispectrum, a detection of a squeezed-limit tensor-scalar-scalar bispectrum through the measurements described in [16, 17], at a level larger than expected from projection effects, would rule out single-field inflation.

The framework we have developed in this paper is also useful in other classes of models that have not yet been directly related to observations. One example is resonant non-Gaussianity [40–44], which is a generic prediction of many models of axion inflation (see [45] and references therein), in particular inflation from axion monodromy [46–48]. In these models the inflationary potential has sinusoidal modulations, which lead to oscillations in the time evolution of the background. Since these oscillations average to zero over a period, they can give corrections to the slow-roll parameter that are larger than is usually allowed and nevertheless be perfectly compatible with current power spectrum data (see e.g. [47, 49, 50]). The resonant bispectrum satisfies Maldacena’s consistency condition [42, 51], but because of oscillations the amplitude of the primordial bispectrum in the squeezed limit is not suppressed by small slow-roll corrections. In light of this, one might wonder whether these models lead to some detectable signal in the scale dependent halo bias. Cyr-Racine and Schmidt [52] showed that indeed resonant non-Gaussianity produces oscillations in the mass dependence of the non-Gaussian halo bias, which is a very unique signature. However, the leading contribution in the large-scale limit to the effect derived in that paper is only a coordinate artefact as shown here. While the halo bias thus has to asymptote to a scale-independent value in the large-scale limit, we expect some interesting effects on intermediate scales. To compute the actual size of this effect it is important to re-write the primordial resonant bispectrum in terms of FNC. It would be very interesting to perform this analysis using the framework constructed in this paper. There are also other models of the early universe (both inflationary and not) in which the metric perturbations do not freeze outside of the horizon. For example, it has been argued in [19, 20] (see also [53]) that this allows one to violate Maldacena’s consistency condition. Another interesting model with a peculiar behavior in the squeezed limit is Khronon inflation, studied in [54]. It would be interesting to use our approach to derive the observational prediction of these and similar models.

On the other hand, multifield inflationary models in general feature a non-trivial bispectrum in the squeezed limit. In this case, curvature perturbations evolve outside the horizon, and the small-scale fluctuations are sensitive to their presence as they essentially evolve in a different FRW background (or, more generally, in a homogeneous anisotropic Universe). This reasoning can be used to connect the squeezed-limit correlators in multifield inflation to late-time observables in a similar way as outlined here for single-field inflation.

Acknowledgments

We acknowledge useful discussions with Liang Dai, Raphael Flauger, Steven Giddings, Donghui Jeong, Marc Kamionkowski, and Justin Khoury. E. P. is supported in part by the Department of Energy grant DE-FG02-91ER-40671. F. S. is supported by NASA through Einstein Postdoctoral Fellowship grant number PF2-130100 awarded by the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for NASA under contract NAS8-03060. M. Z. is supported in part by the National Science Foundation grants PHY-0855425, AST-0907969, PHY-1213563 and by the David & Lucile Packard foundation.

Appendix A: Coordinate transformation of two-point correlations

Consider a scalar field $\rho(x) = \rho(\mathbf{x}, \tau)$. Under a general coordinate transformation $x \rightarrow x'(x)$, ρ transforms as

$$\rho'(x') = \rho(x(x')). \quad (\text{A1})$$

We define the unequal-time two-point correlation of ρ , measured within a patch p through

$$\begin{aligned} \langle \rho(\mathbf{x}, \tau_1) \rho(\mathbf{x} + \mathbf{r}, \tau_2) \rangle_p &= \langle [\bar{\rho}(\tau_1) + \delta\rho(\mathbf{x}, \tau_1)] [\bar{\rho}(\tau_2) + \delta\rho(\mathbf{x} + \mathbf{r}, \tau_2)] \rangle_p \\ &= \bar{\rho}(\tau_1) \bar{\rho}(\tau_2) + \xi(\mathbf{r}, \tau_1, \tau_2), \end{aligned} \quad (\text{A2})$$

where a subscript p indicates an average over the patch, and we have defined $\bar{\rho}(\tau) = \langle \rho(\mathbf{x}, \tau) \rangle_p$, so that by construction the average over the patch of $\delta\rho(\mathbf{x}, \tau)$ vanishes. The correlation function introduced above is then defined as

$$\xi(\mathbf{r}, \tau_1, \tau_2) \equiv \langle \delta\rho(\mathbf{x}, \tau_1) \delta\rho(\mathbf{x} + \mathbf{r}, \tau_2) \rangle_p. \quad (\text{A3})$$

We will consider an alternative definition, the correlation function of $\delta\rho/\rho$, below and show that it transforms in the same way. Further, we will denote the equal time correlation function as $\xi(\mathbf{r}, \tau)$ in the following for simplicity.

Now let us consider the equal-time two-point correlation for ρ' , measured in the same patch:

$$\langle \rho'(\mathbf{x}'_1, \tau') \rho'(\mathbf{x}'_2, \tau') \rangle_p = \langle \rho(\mathbf{x}_1, \tau_1) \rho(\mathbf{x}_2, \tau_2) \rangle_p, \quad (\text{A4})$$

where $\mathbf{x}'_2 - \mathbf{x}'_1 = \mathbf{r}'$, and we have defined for convenience $\mathbf{x}_1 \equiv \mathbf{x}(\mathbf{x}'_1, \tau')$, $\mathbf{x}_2 \equiv \mathbf{x}(\mathbf{x}'_2, \tau')$, and analogously for τ_1, τ_2 . By assumption, the coordinate transform is slowly varying over the patch. We then Taylor expand the coordinate transformations around the center point $\mathbf{x}'_0 = (\mathbf{x}'_1 + \mathbf{x}'_2)/2$,

$$\begin{aligned} x_1^i &= x_0^i - \frac{1}{2} \frac{\partial x^i(\mathbf{x}'_0, \tau')}{\partial x'^j} r'^j; & x_2^i &= x_0^i + \frac{1}{2} \frac{\partial x^i(\mathbf{x}'_0, \tau')}{\partial x'^j} r'^j \\ \tau_1 &= \tau_0 - \frac{1}{2} \frac{\partial \tau(\mathbf{x}'_0, \tau')}{\partial x^j} r'^j; & \tau_2 &= \tau_0 + \frac{1}{2} \frac{\partial \tau(\mathbf{x}'_0, \tau')}{\partial x^j} r'^j, \end{aligned} \quad (\text{A5})$$

where $\mathbf{x}_0 = \mathbf{x}(\mathbf{x}'_0, \tau')$, and $\tau_0 = \tau(\mathbf{x}'_0, \tau')$. We will discuss below why going to linear order in r' is sufficient. We then have

$$\begin{aligned} & \langle \rho'(\mathbf{x}'_1, \tau') \rho'(\mathbf{x}'_2, \tau') \rangle_p \\ &= \bar{\rho}(\tau_1) \bar{\rho}(\tau_2) + \xi \left(\frac{\partial x^i(\mathbf{x}'_0, \tau')}{\partial x'^j} r'^j, \tau_1, \tau_2 \right) \\ &= \left[1 + \frac{1}{2} \frac{\partial \tau(\mathbf{x}'_0, \tau')}{\partial x^j} r'^j \frac{\partial}{\partial \tau_1} - \frac{1}{2} \frac{\partial \tau(\mathbf{x}'_0, \tau')}{\partial x^j} r'^j \frac{\partial}{\partial \tau_2} \right] \\ & \quad \times \left[\bar{\rho}(\tau_1) \bar{\rho}(\tau_2) + \xi \left(\frac{\partial x^i(\mathbf{x}'_0, \tau')}{\partial x'^j} r'^j, \tau_1, \tau_2 \right) \right]_{\tau_1=\tau_2=\tau_0} \\ &= \bar{\rho}^2(\tau_0) + \xi \left(\frac{\partial x^i(\mathbf{x}'_0, \tau')}{\partial x'^j} r'^j, \tau_0 \right), \end{aligned} \quad (\text{A6})$$

since $\xi(\mathbf{r}, \tau_1, \tau_2)$ is symmetric in τ_1 and τ_2 . Thus, at linear order in r' , the equal time correlator in primed coordinates is an equal time correlator in unprimed coordinates. We thus have

$$\begin{aligned} \langle \rho'(\mathbf{x}', \tau') \rho'(\mathbf{x}' + \mathbf{r}', \tau') \rangle_p &= \bar{\rho}^2(\tau_0) + \xi(A\mathbf{r}', \tau_0) \\ A^i_j &= \frac{\partial x^i(\mathbf{x}'_0, \tau')}{\partial x'^j}. \end{aligned} \quad (\text{A7})$$

We now define the correlation function of ρ' in the same way as that for ρ [Eq. (A2)], which yields

$$\langle \rho'(\mathbf{x}', \tau') \rho'(\mathbf{x}' + \mathbf{r}', \tau') \rangle_p = \bar{\rho}'^2(\tau') + \xi'(\mathbf{r}', \tau'), \quad (\text{A8})$$

where

$$\bar{\rho}'(\tau') = \langle \rho'(\mathbf{x}', \tau') \rangle_p = \langle \rho(x(\mathbf{x}', \tau')) \rangle_p = \bar{\rho}(\tau_0). \quad (\text{A9})$$

The second equality again holds to linear order in $\mathbf{x}' - \mathbf{x}'_0$, at which order the average of τ over the patch at fixed τ' is just τ_0 . We will discuss this definition of $\bar{\rho}'$ in App. A 1 below. Thus, the first term in each of Eqs. (A7)–(A8) agrees, and hence so must the second term:

$$\xi'(\mathbf{r}', \tau') = \xi(\mathbf{x}_2 - \mathbf{x}_1, \tau_0) = \xi(A\mathbf{r}', \tau(\mathbf{x}'_0, \tau')). \quad (\text{A10})$$

We can similarly derive what happens to $\tilde{\xi} = \xi/\bar{\rho}^2$, which is the correlation function defined in terms of $\delta\rho/\bar{\rho}$. In this case, we have

$$\langle \rho(\mathbf{x}, \tau) \rho(\mathbf{x} + \mathbf{r}, \tau) \rangle = \bar{\rho}^2(\tau) \left[1 + \tilde{\xi}(\mathbf{r}, \tau) \right]. \quad (\text{A11})$$

For ρ' , this yields

$$\langle \rho'(\mathbf{x}', \tau') \rho'(\mathbf{x}' + \mathbf{r}', \tau') \rangle_p = \bar{\rho}'^2(\tau_0) \left[1 + \tilde{\xi}(\mathbf{x}_2 - \mathbf{x}_1, \tau_0) \right]. \quad (\text{A12})$$

Now we define the correlation function of ρ' in the same way as that for ρ [Eq. (A2)], which yields

$$\langle \rho'(\mathbf{x}', \tau') \rho'(\mathbf{x}' + \mathbf{r}', \tau') \rangle_p = \bar{\rho}'^2(\tau') \left[1 + \tilde{\xi}'(\mathbf{r}', \tau') \right]. \quad (\text{A13})$$

Thus, we see that $\tilde{\xi}$ transforms in the same way as ξ , namely

$$\tilde{\xi}'(\mathbf{r}', \tau') = \tilde{\xi}(\mathbf{x}_2 - \mathbf{x}_1, \tau_0) = \tilde{\xi}(A\mathbf{r}', \tau(\mathbf{x}'_0, \tau')). \quad (\text{A14})$$

There are thus two ways in which the general affine coordinate transform affects the correlation function in primed coordinates: first, there is the spatial transformation of the separation vector, $\mathbf{r} \rightarrow A\mathbf{r}'$; second, the primed correlation function is evaluated at a different point in space and time (here, we have only made the time shift explicit in the notation, but note that the correlation function on the right hand side is to be evaluated within a patch centered on $\mathbf{x}(\mathbf{x}'_0, \tau')$).

The linear order expansion in Eq. (A5) neglects higher derivative terms in the coordinate transformation, i.e. it is valid in the limit that the transformation is slowly varying over the patch. The terms we are neglecting in Eq. (A5) correspond to terms of order k_L^2 and higher, where k_L is the wavenumber of the long-wavelength mode that contributes to the coordinate transformation. Since we are neglecting terms of order k_L^2 throughout this paper, it is sufficient to expand to linear order.

A second simplification occurs because we are only considering three-point correlations in this paper. In this case, it is sufficient to consider the linear response of $\xi'(\mathbf{r}', \tau')$ to the coordinate transformation. We thus write

$$\begin{aligned} A^i_j &= \delta^i_j - a^i_j \\ \Delta\tau &= \tau(\mathbf{x}'_0, \tau') - \tau' \end{aligned} \quad (\text{A15})$$

and work to linear order in a^i_j and $\Delta\tau$. Further, we can neglect the effect of the spatial shift in the position at which ξ is evaluated. This is because the contribution from this shift,

$$\xi'(\mathbf{r}', \tau') \supset [\mathbf{x}(\mathbf{x}'_0, \tau') - \mathbf{x}'_0] \cdot \nabla_{\mathbf{x}'_0} \xi(\mathbf{r}', \tau')|_{\mathbf{x}'_0} \quad (\text{A16})$$

is second order, since both the spatial shift and the location dependence of ξ are at least linear order in the long-wavelength perturbations. Eq. (A10) then simplifies to Eq. (4),

$$\xi'(\mathbf{r}'; \tau') = \left[1 - a_{ij} r'^i \partial_{r'}^j + \Delta\tau \partial_{\tau'} \right] \xi(\mathbf{r}'; \tau'). \quad (\text{A17})$$

1. Coordinate transformations in the presence of non-trivial backgrounds

There is a conceptual point to clarify about the transformation properties of perturbations under changes of coordinates. There are two different but equivalent ways to work with coordinate transformations when studying perturbations on top of some non-trivial background. The first one consists in changing every field in the same way, whether it is going to be treated as a background or not in the rest of the computation. This “democratic” approach is conceptually very simple. For example, if a scalar field transforms as $\rho'(x') = \rho(x(x'))$, so does its

background $\bar{\rho}'(x') = \langle \rho'(x') \rangle = \langle \rho(x) \rangle = \bar{\rho}(x)$. Using this point of view, perturbation of the scalar field around that same background must transform in the same way as well. To see this, consider $\delta(x) \equiv \rho(x) - \bar{\rho}(x)$ (notice that $\langle \delta \rangle = 0$). Then

$$\delta'(x') = \rho'(x') - \bar{\rho}'(t') = \rho(x) - \bar{\rho}(t) = \delta(x), \quad (\text{A18})$$

In the second approach to the problem one splits the field in perturbation and background in such a way that the latter is invariant under coordinate transformations, i.e. $\bar{\rho}'(x) = \bar{\rho}(x)$ (in the cosmological context, the background quantity is only a function of τ). This simplification comes at the cost of more complicated transformation laws for the perturbations, which now do *not* transform as a scalar field. Considering the same example as above and a transformation $x' = x + \epsilon$, at linear order in ϵ one gets

$$\begin{aligned} \delta'(x') &= \rho'(x') - \bar{\rho}'(x') = \rho'(x') - \bar{\rho}(x') \\ &= \rho(x) - \bar{\rho}(x) + \epsilon^\mu \partial_\mu \bar{\rho}(x) = \delta(x) + \epsilon^\mu \partial_\mu \bar{\rho}(x). \end{aligned} \quad (\text{A19})$$

Notice that in both approaches $\rho'(x') = \rho(x')$, while the two different conventions above concern only the

transformations of background and perturbations. A similar discussion can be given for tensor fields such as the metric. Again one has two choices: either one defines metric perturbations that transform like the component of a tensor, in which case the metric background changes when one changes coordinates or the background is assumed fixed in any coordinates and metric perturbations have additional terms in their transformations.

We warn the reader that in the derivation of App. A we have used the “democratic convention” as opposed to the fixed background convention which is standard in the cosmological literature (e.g. [55]). However, it should be kept in mind that around a homogeneous background with a long wavelength perturbation, the additional contribution in the last term of Eq. (A19) is just a constant or a pure gradient at the order we are working in, and hence does not contribute to the correlation function evaluated on much smaller scales r . This means that our final result Eq. (A17) is valid independently of the convention used for defining the perturbations.

Appendix B: Squeezed-limit three-point function from transformed two-point correlation

This section derives the squeezed-limit three-point function and bispectrum from a coordinate transformation of the two-point function given by Eq. (A17),

$$\xi'(\mathbf{r}; \tau) = [1 - a_{ij} r^i \partial_r^j + \Delta \tau \partial_\tau] \xi(\mathbf{r}; \tau), \quad (\text{B1})$$

where we have dropped the prime on coordinates since we will only deal with primed coordinates in this section. Specifically, we want to derive the three point function $\langle X'(\mathbf{x}_3) \delta'(\mathbf{x}_1) \delta'(\mathbf{x}_2) \rangle$ in the limit where $|\mathbf{x}_3 - \mathbf{x}_1| \gg |\mathbf{x}_1 - \mathbf{x}_2|$ (squeezed limit). Here, X and δ stand for any perturbation variables with mean zero (of course, the auto-three-point function $X = \delta$ is a special case). In this limit, the three-point function quantifies the modulation of the local two-point function $\xi'(|\mathbf{x}_1 - \mathbf{x}_2|)$ by a long-wavelength perturbation X' (modes with wavelength much less than $|\mathbf{x}_3 - \mathbf{x}_1|$ will not contribute to this correlation). As discussed in Sec. II, the coordinate transformation only acts on the small-scale fluctuations, so that we set $X'(\mathbf{x}) = X(\mathbf{x})$. Thus,

$$\langle X'(\mathbf{x}_3) \delta'(\mathbf{x}_1) \delta'(\mathbf{x}_2) \rangle \stackrel{\text{squeezed}}{=} \langle X'(\mathbf{x}_3) \xi'(\mathbf{x}_1 - \mathbf{x}_2; \tau) |_{X'(\mathbf{x}_c)} \rangle, \quad (\text{B2})$$

where we have defined the point of evaluation of the long-wavelength X perturbation as

$$\mathbf{x}_c = c\mathbf{x}_1 + (1 - c)\mathbf{x}_2. \quad (\text{B3})$$

Choosing \mathbf{x}_c along the axis connecting \mathbf{x}_1 and \mathbf{x}_2 is necessary since Eq. (B2) describes a homogeneous and isotropic three-point function. The midpoint $\mathbf{x}_0 = (\mathbf{x}_1 + \mathbf{x}_2)/2$ corresponds to $c = 1/2$. We will see that the choice of c does not influence the final result below. Using Eq. (B1), this yields

$$\begin{aligned} \langle X'(\mathbf{x}_3) \delta'(\mathbf{x}_1) \delta'(\mathbf{x}_2) \rangle \stackrel{\text{squeezed}}{=} & [-\xi_{X a_{ij}}(|\mathbf{x}_3 - \mathbf{x}_c|) r^i \partial_r^j + \xi_{X \Delta \tau}(|\mathbf{x}_3 - \mathbf{x}_c|) \partial_\tau] \xi(\mathbf{x}_1 - \mathbf{x}_2; \tau) + \langle X(\mathbf{x}_3) \xi(\mathbf{x}_1 - \mathbf{x}_2; \tau) |_{X(\mathbf{x}_c)} \rangle \\ & = [-\xi_{X a_{ij}}(|\mathbf{x}_3 - \mathbf{x}_c|) r^i \partial_r^j + \xi_{X \Delta \tau}(|\mathbf{x}_3 - \mathbf{x}_c|) \partial_\tau] \xi(\mathbf{x}_1 - \mathbf{x}_2; \tau) + \langle X(\mathbf{x}_3) \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \rangle. \end{aligned} \quad (\text{B4})$$

This expression gives the squeezed-limit three-point function of X' and δ' (in the *primed* coordinates) in terms of derivatives of $\xi(\mathbf{r}, \tau)$ and the three-point function of X and δ in the *unprimed* coordinate frame.

The left-hand side of Eq. (B4) is clearly symmetric under $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$. For a general choice of \mathbf{x}_c , this does not hold

for the right-hand side, so we should symmetrize:

$$\begin{aligned} \langle X'(\mathbf{x}_3)\delta'(\mathbf{x}_1)\delta'(\mathbf{x}_2) \rangle \stackrel{\text{squeezed}}{=} \frac{1}{2} \left[-\xi_{Xa_{ij}}(|\mathbf{x}_3 - \mathbf{x}_c|) r^i \partial_r^j - \xi_{Xa_{ij}}(|\mathbf{x}_3 - \mathbf{x}'_c|) r^i \partial_r^j \right. \\ \left. + \xi_{X\Delta\tau}(|\mathbf{x}_3 - \mathbf{x}_c|) \partial_\tau + \xi_{X\Delta\tau}(|\mathbf{x}_3 - \mathbf{x}'_c|) \partial_\tau \right] \xi(\mathbf{x}_1 - \mathbf{x}_2; \tau) + \langle X(\mathbf{x}_3)\delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle, \end{aligned} \quad (\text{B5})$$

where $\mathbf{x}'_c = (1 - c)\mathbf{x}_1 + c\mathbf{x}_2$. Of course, if we choose \mathbf{x}_c to be the midpoint of \mathbf{x}_1 and \mathbf{x}_2 , the two permutations are identical.

We now derive the Fourier-space analog of Eq. (B5). In terms of the cross-power spectra $P_{Xa_{ij}}(k)$, $P_{X\Delta\tau}(k)$ and the auto power spectrum $P_\delta(k)$, we have

$$\begin{aligned} \langle X(\mathbf{x}_3)a_{ij}(\mathbf{x}_c) \rangle &= \int \frac{d^3\mathbf{k}_L}{(2\pi)^3} P_{Xa_{ij}}(k_L) e^{i\mathbf{k}_L[\mathbf{x}_3 - \mathbf{x}_c]} \\ \langle X(\mathbf{x}_3)\Delta\tau(\mathbf{x}_c) \rangle &= \int \frac{d^3\mathbf{k}_L}{(2\pi)^3} P_{X\Delta\tau}(k_L) e^{i\mathbf{k}_L[\mathbf{x}_3 - \mathbf{x}_c]} \\ \bar{r}^j \partial_i \xi(\mathbf{r}, \tau) &= \int \frac{d^3\mathbf{k}_S}{(2\pi)^3} P_\delta(k_S, \tau) \bar{r}^j \partial_i e^{i\mathbf{k}_S \mathbf{r}} \\ &= - \int \frac{d^3\mathbf{k}_S}{(2\pi)^3} \frac{\partial}{\partial k_S^j} [k_S^i P_\delta(k_S)] e^{i\mathbf{k}_S \mathbf{r}} \\ \partial_\tau \xi(\bar{\mathbf{r}}, \tau) &= \int \frac{d^3\mathbf{k}_S}{(2\pi)^3} \partial_\tau P_\delta(k_S, \tau) e^{i\mathbf{k}_S \mathbf{r}}. \end{aligned} \quad (\text{B6})$$

The bispectrum in the squeezed limit $k_L \ll k_1, k_2$ becomes

$$\begin{aligned} \langle X'(\mathbf{k}_L)\delta'(\mathbf{k}_1)\delta'(\mathbf{k}_2) \rangle &= \int d^3\mathbf{x}_3 \int d^3\mathbf{x}_1 \int d^3\mathbf{x}_2 e^{-i(\mathbf{k}_L\mathbf{x}_3 + \mathbf{k}_1\mathbf{x}_1 + \mathbf{k}_2\mathbf{x}_2)} \langle X(\mathbf{x}_3)\delta'(\mathbf{x}_1)\delta'(\mathbf{x}_2) \rangle \\ &\stackrel{\text{squeezed}}{=} \frac{1}{2} \left\{ \left[P_{Xa_{ij}}(k_L) \frac{\partial}{\partial k_S^j} [k_S^i P_\delta(k_S)] + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\delta(k_S) \right]_{\mathbf{k}_S = \mathbf{k}_1 + c\mathbf{k}_L} + (c \rightarrow 1 - c) \right\} \\ &\quad \times (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_L) + \langle X(\mathbf{k}_L)\delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle. \end{aligned} \quad (\text{B7})$$

Thus, the transformed bispectrum has the proper delta function ensuring the triangle condition, and we can identify the $X'\delta'\delta'$ -bispectrum in the squeezed limit as

$$\begin{aligned} B_{X'\delta'\delta'}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) &\stackrel{\text{squeezed}}{=} \frac{1}{2} \left\{ \left[P_{Xa_{ij}}(k_L) \frac{\partial}{\partial k_S^j} [k_S^i P_\delta(k_S)] + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\delta(k_S) \right]_{\mathbf{k}_S = \mathbf{k}_1 + c\mathbf{k}_L} + (c \rightarrow 1 - c) \right\} \\ &\quad + B_{X\delta\delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) \\ &= \frac{1}{2} \left\{ \left[P_{Xa_{ij}}(k_L) \left(\delta_{ij} + \frac{k_S^i k_S^j}{k_S^2} \frac{d}{d \ln k_S} \right) P_\delta(k_S) + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\delta(k_S) \right]_{\mathbf{k}_S = \mathbf{k}_1 + c\mathbf{k}_L} + (c \rightarrow 1 - c) \right\} \\ &\quad + B_{X\delta\delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \end{aligned} \quad (\text{B8})$$

Since we are working in the squeezed limit, we can expand in k_1/k_L . We have

$$\begin{aligned} F(|\mathbf{k}_1 + c\mathbf{k}_L|) &= F(k_1) + c\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_L \frac{k_L}{k_1} \frac{d}{d \ln k_1} F(k_1) + \mathcal{O}(k_L^2/k_1^2) \\ \frac{k_S^i k_S^j}{k_S^2} &= \frac{k_1^i k_1^j}{k_1^2} \left(1 - 2c \frac{\mathbf{k}_1 \cdot \mathbf{k}_L}{k_1^2} \right) + 2c \frac{k_1^i k_L^j}{k_1^2} + \mathcal{O}(k_L^2/k_1^2) \\ \frac{dP_\delta(k_S)}{d \ln k_S} &= \frac{dP_\delta(k_1)}{d \ln k_1} + c\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_L \frac{k_L}{k_1} \frac{d^2}{d(\ln k_1)^2} P_\delta(k_1) + \mathcal{O}(k_L^2/k_1^2). \end{aligned} \quad (\text{B9})$$

We already see that all corrections are linear in c , so that the two permutations $c, 1 - c$ add up to 1. Thus, the result becomes independent of the choice of c , and the squeezed limit bispectrum in terms is, to order (k_L^2/k_1^2) ,

$$\begin{aligned}
B_{X'\delta'\delta'}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) \stackrel{\text{squeezed}}{=} & P_{Xa_{ij}}(k_L) \left(\delta_{ij} + \frac{k_1^i k_1^j}{k_1^2} \frac{d}{d \ln k_1} \right) P_\delta(k_1) + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\delta(k_1) \\
& + \hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_L \frac{k_L}{k_1} \frac{1}{2} \left\{ P_{Xa_{ij}}(k_L) \left(\delta_{ij} + \frac{k_1^i k_1^j}{k_1^2} \left[\frac{d}{d \ln k_1} - 2 \right] \right) \frac{dP_\delta(k_1)}{d \ln k_1} + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} \frac{dP_\delta(k_1)}{d \ln k_1} \right\} \\
& + P_{Xa_{ij}}(k_L) \frac{k_1^i k_L^j}{k_1^2} \frac{dP_\delta(k_1)}{d \ln k_1} + B_{X\delta\delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \tag{B10}
\end{aligned}$$

Since $\mathbf{k}_2 = -\mathbf{k}_1 + \mathcal{O}(k_L/k_1)$, this expression can be equivalently written in terms of \mathbf{k}_2 instead of \mathbf{k}_1 . Since the choice of c is arbitrary, we will use the most natural choice, $c = 1/2$, in which case Eq. (B8) becomes simply

$$\begin{aligned}
B_{X'\delta'\delta'}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) \stackrel{\text{squeezed}}{=} & \left[P_{Xa_{ij}}(k_L) \left(\delta_{ij} + \frac{k_S^i k_S^j}{k_S^2} \frac{d}{d \ln k_S} \right) P_\delta(k_S) + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\delta(k_S) \right]_{\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2} \\
& + B_{X\delta\delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \tag{B11}
\end{aligned}$$

While this expression is equivalent to Eq. (B10) up to order k_L^2 , we will work with this result since it is more compact and convenient.

Including the effect of a mean density modulated by X following Eq. (44) in Sec. IV is now an obvious generalization. We obtain

$$\begin{aligned}
B_{X'\delta'\delta'}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) \stackrel{\text{squeezed}}{=} & \left[P_{Xa_{ij}}(k_L) \left(\delta_{ij} + \frac{k_S^i k_S^j}{k_S^2} \frac{d}{d \ln k_S} \right) P_\delta(k_S) + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\delta(k_S) + 2P_{Xc}(k_L) P_\delta(k_S) \right]_{\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2} \\
& + B_{X\delta\delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \tag{B12}
\end{aligned}$$

We can now further decompose a_{ij} as

$$a_{ij} = \frac{1}{n_D} a \delta_{ij} + a_{ij}^T, \tag{B13}$$

where n_D is the dimensionality of the space in which we define the correlations, and a_{ij}^T is traceless. In particular $n_D = 3$ for three-dimensional observables such as galaxy densities or 21cm flux, and $n_D = 2$ for projected quantities on the sky such as the CMB. We then obtain

$$\begin{aligned}
B_{X'\delta'\delta'}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) \stackrel{\text{squeezed}}{=} & \left[P_{Xa}(k_L) \left(1 + \frac{1}{n_D} \frac{d}{d \ln k_S} \right) P_\delta(k_S) + P_{Xa_{ij}^T}(k_L) \frac{k_S^i k_S^j}{k_S^2} \frac{dP_\delta(k_S)}{d \ln k_S} \right. \\
& \left. + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\delta(k_S) + 2P_{Xc}(k_L) P_\delta(k_S) \right]_{\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2} + B_{X\delta\delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) \\
= & \left[P_{Xa}(k_L) P_\delta(k_S) \frac{d \ln(k_S^{n_D} P_\delta(k_S))}{d \ln k_S} + P_{Xa_{ij}^T}(k_L) \frac{k_S^i k_S^j}{k_S^2} \frac{dP_\delta(k_S)}{d \ln k_S} \right. \\
& \left. + P_{X\Delta\tau}(k_L) \frac{\partial}{\partial \tau} P_\delta(k_S) + 2P_{Xc}(k_L) P_\delta(k_S) \right]_{\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2} + B_{X\delta\delta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2). \tag{B14}
\end{aligned}$$

Note that the contribution of the trace of a_{ij} scales as the logarithmic derivative of $k^{n_D} P_\delta(k)$, while the trace-free component a_{ij}^T couples to the logarithmic derivative of $P_\delta(k)$ itself.

Appendix C: Fermi Normal Coordinates

do lead to a metric of the form

1. FNC construction and metric

In this section, we show how for a general metric, the coordinates x_F^μ defined through Eq. (C2) below in fact

$$g_{\mu\nu}^F = \eta_{\mu\nu} + \mathcal{O}(x_F^2). \tag{C1}$$

To quadratic order, the Fermi coordinate transformation is given by

$$x^\mu(x_F^i)\Big|_{t_P} = x^\mu(P) + (e_i)_P^\mu x_F^i - \frac{1}{2}\Gamma_{\alpha\beta}^\mu\Big|_P (e_i)_P^\alpha (e_j)_P^\beta x_F^i x_F^j. \quad (\text{C2})$$

Straightforward algebra yields the transformation matrix as

$$\frac{\partial x^\alpha}{\partial x_F^\mu} = (e_\mu)_P^\alpha + A_{\mu j}^\alpha x_F^j + \mathcal{O}(x_F^2), \quad (\text{C3})$$

where a subscript P denotes the evaluation at the point on the central geodesic specified by x_F^0 . Here,

$$A_{0j}^\alpha = (e_0)_P^\beta [\partial_\beta (e_j)^\alpha]_P \\ A_{ij}^\alpha = -\Gamma_{\beta\gamma}^\alpha\Big|_P (e_i)_P^\beta (e_j)_P^\gamma \quad (\text{C4})$$

Using the fact that the unit vectors $(e_j)^\alpha$ are assumed to be parallel-transported along the central geodesic, we have

$$0 = (e_0)_P^\beta [\nabla_\beta (e_j)^\alpha]_P = (e_0)_P^\beta \{\partial_\beta (e_j)^\alpha + \Gamma_{\beta\lambda}^\alpha (e_j)^\lambda\}_P,$$

which can be used to bring $A_{\mu j}^\alpha$ into a uniform expression:

$$A_{\mu j}^\alpha = -\Gamma_{\beta\gamma}^\alpha\Big|_P (e_\mu)_P^\beta (e_j)_P^\gamma. \quad (\text{C5})$$

Next, we expand

$$g_{\alpha\beta}(x(x_F)) = g_{\alpha\beta}\Big|_P + (e_j)_P^\gamma [\partial_\gamma g_{\alpha\beta}]_P x_F^j. \quad (\text{C6})$$

Finally, using that

$$g_{\alpha\beta}\Big|_P (e_\mu)_P^\alpha (e_\nu)_P^\beta = \eta_{\mu\nu} \quad (\text{C7})$$

by construction of the orthonormal tetrad (note that this holds for any velocity v_i at order v), we obtain

$$g_{\mu\nu}^F(x_F) = g_{\alpha\beta}(x(x_F)) \frac{\partial x^\alpha}{\partial x_F^\mu} \frac{\partial x^\beta}{\partial x_F^\nu} \\ = \eta_{\mu\nu} + [\partial_\gamma g_{\alpha\beta} - 2g_{\lambda\beta}\Gamma_{\alpha\gamma}^\lambda] (e_\mu)_P^\alpha (e_\nu)_P^\beta (e_j)_P^\gamma x_F^j \\ + \mathcal{O}(x_F^2). \quad (\text{C8})$$

Assuming that we have a metric connection so that

$$\nabla_\gamma g_{\alpha\beta} = \partial_\gamma g_{\alpha\beta} - 2\Gamma_{\gamma(\alpha}^\lambda g_{\lambda\beta)} = 0, \quad (\text{C9})$$

we see that the order x_F^j correction to the metric vanishes.

So far, we have not used the condition that the central curve whose tangent vector is $(e_0)^\mu$ is a geodesic. However, if we want the metric in FNC to be of the form Eq. (C1) at more than just one point along the central curve, we clearly need at lowest order

$$(e_0)^\gamma \nabla_\gamma [g_{\alpha\beta} (e_\mu)^\alpha (e_\nu)^\beta]_P = 0. \quad (\text{C10})$$

This in particular implies

$$(e_0)^\gamma \nabla_\gamma (e_0)^\mu = 0 \quad (\text{C11})$$

along the curve, which is precisely the condition for the central curve to be a geodesic.

2. Transformation from general coordinates to $\overline{\text{FNC}}$

We now explicitly derive the transformation Eqs. (12)–(13) into the $\overline{\text{FNC}}$ frame. We write the conformal metric as

$$\bar{g}_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (\text{C12})$$

and work to linear order in $h_{\mu\nu}$. The tetrad around P is given by

$$(e_0)^\mu = \left(1 + \frac{1}{2}h_{00}, v^i\right) \\ (e_j)^\mu = \left(v_j + h_{0j}, \delta_j^i - \frac{1}{2}h_j^i\right). \quad (\text{C13})$$

Note that we treat v as $\mathcal{O}(h)$ here, as we are assuming comoving observers. Since $(e_\alpha)^\mu = \delta_\alpha^\mu + \mathcal{O}(h)$ and $\Gamma_{\alpha\beta}^\mu = \mathcal{O}(h)$, we have for the transformation into $\overline{\text{FNC}}$ at lowest order in h :

$$x^\mu(\bar{x}_F^\alpha) = P^\mu + (e_i)_P^\mu \bar{x}_F^i - \frac{1}{2}\Gamma_{ij}^\mu\Big|_P x_F^i x_F^j, \quad (\text{C14})$$

where $P^\mu \equiv x^\mu(P)$ are the global coordinates of the central geodesic, so that $v^i = \partial P^i / \partial x^0$ and we define analogously $\bar{P}^\mu \equiv \bar{x}_F^\mu(P)$ for $\overline{\text{FNC}}$. The conformal proper time \bar{P}^0 defines the time coordinate of $\overline{\text{FNC}}$, namely $\bar{P}^0 = \bar{x}_F^0$. Without loss of generality, we can choose the spatial origin so that $\bar{P}^i(\tau_F) = 0$ at some fixed proper time τ_F . To lowest order in h and v , the conformal proper time is related to the global time by

$$x^0(P) = \bar{x}_F^0 + \frac{1}{2} \int_0^{\bar{x}_F^0} h_{00}(\mathbf{0}, \tau) d\tau. \quad (\text{C15})$$

The Christoffel symbols are given by

$$\Gamma_{ij}^\mu = \frac{1}{2} [h_{i,j}^\mu + h_{j,i}^\mu - h_{ij}^{\prime\mu}], \quad (\text{C16})$$

where indices are raised and lowered with $\eta_{\mu\nu}$. Thus,

$$x^0(\bar{x}_F^\alpha) = \bar{x}_F^0 + \frac{1}{2} \int_0^{\bar{x}_F^0} h_{00}(\tau) d\tau + (v_i + h_{0i}) \bar{x}_F^i \\ - \frac{1}{4} [h_{i,j}^0 + h_{j,i}^0 - h_{ij}^{\prime 0}] \bar{x}_F^i \bar{x}_F^j \\ = \bar{x}_F^0 + \frac{1}{2} \int_0^{\bar{x}_F^0} h_{00}(\tau) d\tau + (v_i + h_{0i}) \bar{x}_F^i \\ + \frac{1}{4} [h_{0i,j} + h_{0j,i} + h_{ij}^{\prime 0}] \bar{x}_F^i \bar{x}_F^j \quad (\text{C17})$$

$$x^k(\bar{x}_F^\alpha) = P^k - \frac{1}{2} h_{i,j}^k \bar{x}_F^i - \frac{1}{4} [h_{i,j}^k + h_{j,i}^k - h_{ij}^{\prime k}] \bar{x}_F^i \bar{x}_F^j. \quad (\text{C18})$$

Here all terms linear in perturbations, namely $h_{\mu\nu}$ and v^i , should be evaluated along the central geodesic P^μ . For $\bar{x}_F^0 \neq \tau_F$, the spatial position of the geodesic differs from the origin in the x^μ coordinate system by an amount

$\Delta x_F^0 v^i$, which is linear in perturbations (we will shortly find v^i from the geodesic equation). Therefore, up to terms quadratic in metric perturbations, we can evaluate $h_{\mu\nu}$ and v^i at the origin of the FNC $\bar{x}_F^i = \bar{P}^i(x_{F*}^0) \equiv 0$.

At linear order in h , the inverse transformation is very easily derived:

$$\begin{aligned} \bar{x}_F^0(x^\alpha) &= x^0 - \frac{1}{2} \int_0^{x^0} h_{00}(\tau) d\tau - (v_i + h_{0i}) x^i \\ &\quad - \frac{1}{4} [h_{0i,j} + h_{0j,i} + h'_{ij}] x^i x^j \end{aligned} \quad (C19)$$

$$\bar{x}_F^k(x^\alpha) = x^k + \frac{1}{2} h^k_{i} x^i + \frac{1}{4} [h^k_{i,j} + h^k_{j,i} - h_{ij}{}^{,k}] x^i x^j. \quad (C20)$$

Note that when $h_{0i} \neq 0$ there is no solution in which the observer is at rest. We now focus on this case now, where we can set $h_{ij} = h_{00} = 0$, since their effect is simply additive and was considered above. For $h_{i0} \neq 0$, one needs to know the relation between v^i and h_{0i} that is enforced by the geodesic equation for P. Assuming as we did previously that the velocity $v^i = \partial P^i / \partial x^0$ is of the same order as h_{0i} , the geodesic equations at linear order in perturbations in global coordinates are given by

$$\left(\frac{\partial}{\partial x^0} \right)^2 P^0 = 0, \quad \left(\frac{\partial}{\partial x^0} \right)^2 P^i + \Gamma_{00}^i \left(\frac{\partial P^0}{\partial x^0} \right)^2 = 0,$$

where we used the fact that $\Gamma_{\nu\sigma}^\mu$ is at least linear in metric perturbations. The first equation tells us that proper time coincides with the global time coordinate at this order, up to two arbitrary integration constants, which we fix by choosing $P^0 = x^0$. This is no surprise since we are assuming $h_{00} = 0$. Then, from the second equation, we find $v^i = \partial P^i / \partial x^0 = -h_{0i}$, where we have set another integration constant to zero following the assumption that $\partial_0 P^i$ is linear in h_{0i} . Specifically, for \bar{x}_F^0 close to τ_F , we have

$$x^i(P) = v^i(\bar{x}_F^0 - \tau_F), \quad (C21)$$

so that to leading order

$$\begin{aligned} x^k(\bar{x}_F^\alpha) &= v^k(\bar{x}_F^0 - \tau_F) + \bar{x}_F^k - \frac{1}{2} h^k_{i} \bar{x}_F^i \\ &\quad - \frac{1}{4} [h^k_{i,j} + h^k_{j,i} - h_{ij}{}^{,k}] \bar{x}_F^i \bar{x}_F^j \end{aligned} \quad (C22)$$

$$\begin{aligned} \bar{x}_F^k(x^\alpha) &= x^k - v^k(x^0 - \tau_F) + \frac{1}{2} h^k_{i} x^i \\ &\quad + \frac{1}{4} [h^k_{i,j} + h^k_{j,i} - h_{ij}{}^{,k}] x^i x^j. \end{aligned} \quad (C23)$$

Appendix D: Uniqueness of Fermi Normal Coordinates

In this section we investigate what residual coordinate freedom remains when requiring that the metric be of ‘‘FNC form’’:

$$g_{\mu\nu}^F = \eta_{\mu\nu} + S_{\mu\nu ij} x_F^i x_F^j. \quad (D1)$$

Here, $S_{\mu\nu ij}$ in general depends on the affine parameter along the central geodesic, i.e. the Fermi-frame time coordinate t_F . Let us consider a general inertial frame $\{x'^\mu\}$ constructed around point P . Without loss of generality, we let P be at the origin of both the $\{x'^\mu\}$ and $\{x_F^\mu\}$ coordinate systems. The requirement that $\{x'^\mu\}$ be inertial, i.e. that $g'_{\mu\nu} = \eta_{\mu\nu}$ at P and that $\partial'_\alpha g'_{\mu\nu} = 0$ at P , restricts the relation between the coordinates to be at least of cubic order:

$$\begin{aligned} x'^\alpha &= x_F^\alpha + \frac{1}{3} C^\alpha_{\beta\gamma\delta} x_F^\beta x_F^\gamma x_F^\delta + \mathcal{O}(x_F^4) \\ x_F^\alpha &= x'^\alpha - \frac{1}{3} C^\alpha_{\beta\gamma\delta} x'^\beta x'^\gamma x'^\delta + \mathcal{O}(x'^4). \end{aligned} \quad (D2)$$

Here,

$$C^\alpha_{\beta\gamma\delta} = \frac{\partial^3 x'^\alpha}{\partial x_F^\beta \partial x_F^\gamma \partial x_F^\delta} \Big|_P, \quad (D3)$$

which implies

$$C^\alpha_{\beta\gamma\delta} = C^\alpha_{(\beta\gamma)\delta} = C^\alpha_{\beta(\gamma\delta)} = C^\alpha_{(\beta\gamma\delta)}, \quad (D4)$$

i.e. the last equality states that $C^\alpha_{\beta\gamma\delta} = C^\alpha_{\delta\gamma\beta}$. The metric in the primed frame then becomes to order x'^2 :

$$g'_{\mu\nu} = \eta_{\mu\nu} + [S_{\mu\nu\gamma\delta} - 2C_{(\mu\nu)\gamma\delta}] x'^\gamma x'^\delta. \quad (D5)$$

By assumption, $S_{\mu\nu\gamma\delta}$ obeys the ‘‘FNC condition’’

$$S_{\mu\nu 0\delta} = 0 = S_{\mu\nu\gamma 0}. \quad (D6)$$

Imposing the same restriction on $g'_{\mu\nu}$ leads to the condition

$$C_{(\mu\nu)0\delta} = 0 = C_{(\mu\nu)\gamma 0}. \quad (D7)$$

Using the symmetry properties Eq. (D4), we obtain

$$C_{(\mu 0)\gamma\delta} = 0. \quad (D8)$$

This says that there is *no freedom* in the order $\mathcal{O}(x_F^2)$ corrections to the Fermi-frame metric components g_{00}^F and g_{0i}^F for coordinates satisfying the condition Eq. (D1) (but see below). However, there is some freedom in choosing the spatial part g_{ij}^F , since a non-zero C_{ijkl} is allowed by Eq. (D1).

Finally, there is an additional freedom in the choice of coordinates which does not affect the metric at order x_F^2 . Namely, we can choose $C_{[0\mu]ij} \neq 0$. Specifically,

$$C_{0kij} = -C_{k0ij} = 3\alpha_k A_{ij}, \quad (D9)$$

where we have introduced a 3-vector α_k and a symmetric 3-tensor A_{ij} . By construction, $C_{(0\mu)\gamma\delta} = 0$. This corresponds to a coordinate transform of (note that $C^0_{kij} = -\alpha_k A_{ij}$)

$$\begin{aligned} x'^0 &= x_F^0 - \alpha_k x_F^k A_{ij} x_F^i x_F^j \\ x'^i &= x_F^i - \alpha^i x_F^0 A_{lj} x_F^l x_F^j. \end{aligned} \quad (D10)$$

In other words, this coordinate transform is a spatially location-dependent Lorentz boost with a velocity given by

$$v^k(x_F^l) = \alpha^k A_{ij} x_F^i x_F^j, \quad (\text{D11})$$

which vanishes quadratically on the central geodesic. This type of coordinate transform leaves the Fermi frame metric entirely invariant at order x_F^2 .

-
- [1] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, *Phys. Rev. D* **77**, 123514 (2008), 0710.4560.
 - [2] S. Matarrese and L. Verde, *Astrophys. J. Lett.* **677**, L77 (2008), 0801.4826.
 - [3] F. Schmidt and M. Kamionkowski, *Phys. Rev. D* **82**, 103002 (2010), 1008.0638.
 - [4] J. Maldacena, *Journal of High Energy Physics* **5**, 013 (2003), arXiv:astro-ph/0210603.
 - [5] F. K. Manasse and C. W. Misner, *Journal of Mathematical Physics* **4**, 735 (1963).
 - [6] F. Schmidt and D. Jeong, *Phys. Rev. D* **86**, 083527 (2012), 1204.3625.
 - [7] T. Baldauf, U. Seljak, L. Senatore, and M. Zaldarriaga, *JCAP* **10**, 031 (2011), 1106.5507.
 - [8] P. Creminelli, J. Noreña, and M. Simonović, *JCAP* **7**, 052 (2012), 1203.4595.
 - [9] L. Senatore and M. Zaldarriaga (2012), 1210.6048.
 - [10] L. Senatore and M. Zaldarriaga, *JCAP* **1208**, 001 (2012), 1203.6884.
 - [11] P. Creminelli and M. Zaldarriaga, *Phys. Rev. D* **70**, 083532 (2004), arXiv:astro-ph/0405428.
 - [12] L. Boubekeur, P. Creminelli, G. D'Amico, J. Noreña, and F. Vernizzi, *JCAP* **8**, 029 (2009), 0906.0980.
 - [13] P. Creminelli, C. Pitrou, and F. Vernizzi, *JCAP* **11**, 025 (2011), 1109.1822.
 - [14] A. Lewis, *JCAP* **1206**, 023 (2012), 1204.5018.
 - [15] D. Jeong and F. Schmidt, in prep. (2013).
 - [16] K. W. Masui and U.-L. Pen, *Physical Review Letters* **105**, 161302 (2010), 1006.4181.
 - [17] D. Jeong and M. Kamionkowski, *Physical Review Letters* **108**, 251301 (2012), 1203.0302.
 - [18] T. Tanaka and Y. Urakawa, *JCAP* **1105**, 014 (2011), 1103.1251.
 - [19] M. H. Namjoo, H. Firouzjahi, and M. Sasaki, *Europhys. Lett.* **101**, 39001 (2013), 1210.3692.
 - [20] X. Chen, H. Firouzjahi, M. H. Namjoo, and M. Sasaki (2013), 1301.5699.
 - [21] P. Creminelli, G. D'Amico, M. Musso, and J. Noreña, *JCAP* **11**, 038 (2011), 1106.1462.
 - [22] R. Flauger, D. Green, and R. A. Porto, *ArXiv e-prints* (2013), 1303.1430.
 - [23] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, et al., *ArXiv e-prints* (2013), 1303.5079.
 - [24] F. Schmidt and L. Hui, *ArXiv e-prints* (2012), 1210.2965.
 - [25] F. Schmidt, D. Jeong, and V. Desjacques, *ArXiv e-prints* (2012), 1212.0868.
 - [26] F. Schmidt, *ArXiv e-prints* (2013), 1304.1817.
 - [27] V. Desjacques, D. Jeong, and F. Schmidt, *Phys. Rev. D* **84**, 063512 (2011), 1105.3628.
 - [28] R. Scoccimarro, L. Hui, M. Manera, and K. C. Chan, *Phys. Rev. D* **85**, 083002 (2012), 1108.5512.
 - [29] F. Schmidt, *ArXiv e-prints* (2013), 1304.1817.
 - [30] S. Endlich, A. Nicolis, and J. Wang, *ArXiv e-prints* (2012), 1210.0569.
 - [31] J. Yoo, A. L. Fitzpatrick, and M. Zaldarriaga, *Phys. Rev. D* **80**, 083514 (2009), 0907.0707.
 - [32] A. Challinor and A. Lewis, *ArXiv e-prints* (2011), 1105.5292.
 - [33] C. Bonvin and R. Durrer, *ArXiv e-prints* (2011), 1105.5280.
 - [34] D. Jeong, F. Schmidt, and C. M. Hirata, *Phys. Rev. D* **85**, 023504 (2012), 1107.5427.
 - [35] S. B. Giddings and M. S. Sloth, *JCAP* **1**, 023 (2011), 1005.1056.
 - [36] S. B. Giddings and M. S. Sloth, *Phys. Rev. D* **84**, 063528 (2011), 1104.0002.
 - [37] S. Dodelson, E. Rozo, and A. Stebbins, *Physical Review Letters* **91**, 021301 (2003), arXiv:astro-ph/0301177.
 - [38] F. Schmidt and D. Jeong, *Phys. Rev. D* **86**, 083513 (2012), 1205.1514.
 - [39] N. Kaiser and A. Jaffe, *Astrophys. J.* **484**, 545 (1997), arXiv:astro-ph/9609043.
 - [40] X. Chen, R. Easther, and E. A. Lim, *JCAP* **0706**, 023 (2007), astro-ph/0611645.
 - [41] X. Chen, R. Easther, and E. A. Lim, *JCAP* **0804**, 010 (2008), 0801.3295.
 - [42] R. Flauger and E. Pajer, *JCAP* **1101**, 017 (2011), 1002.0833.
 - [43] L. Leblond and E. Pajer, *JCAP* **1101**, 035 (2011), 1010.4565.
 - [44] S. Hannestad, T. Haugbolle, P. R. Jarnhus, and M. S. Sloth, *JCAP* **1006**, 001 (2010), 0912.3527.
 - [45] E. Pajer and M. Peloso (2013), to appear.
 - [46] L. McAllister, E. Silverstein, and A. Westphal, *Phys. Rev. D* **82**, 046003 (2010), 0808.0706.
 - [47] R. Flauger, L. McAllister, E. Pajer, A. Westphal, and G. Xu, *JCAP* **1006**, 009 (2010), 0907.2916.
 - [48] M. Berg, E. Pajer, and S. Sjors, *Phys. Rev. D* **81**, 103535 (2010), 0912.1341.
 - [49] M. Aich, D. K. Hazra, L. Sriramkumar, and T. Souradeep, *Phys. Rev. D* **87**, 083526 (2013), 1106.2798.
 - [50] H. Peiris, R. Easther, and R. Flauger (2013), 1303.2616.
 - [51] P. Creminelli, G. D'Amico, M. Musso, and J. Noreña, *JCAP* **1111**, 038 (2011), 1106.1462.
 - [52] F.-Y. Cyr-Racine and F. Schmidt, *Phys. Rev. D* **84**, 083505 (2011), 1106.2806.
 - [53] W. H. Kinney, *Phys. Rev. D* **72**, 023515 (2005), gr-qc/0503017.
 - [54] P. Creminelli, J. Noreña, M. Pena, and M. Simonovic, *JCAP* **1211**, 032 (2012), 1206.1083.
 - [55] S. Weinberg (2008).